Express Mail No.: EL932931210US
Date of Deposit: December 11, 2003
Atty. Docket No.: 18396/2002B

Method for Generating Diversity

The present invention relates to a method for generating diversity in a gene or gene product by exploiting the natural somatic hypermutation capability of antibody-producing cells, as well as to cell lines capable of generating diversity in defined gene products.

Many *in vitro* approaches to the generation of diversity in gene products rely on the generation of a very large number of mutants which are then selected using powerful selection technologies. For example, phage display technology has been highly successful as providing a vehicle that allows for the selection of a displayed protein (Smith, 1985; Bass *et al.* , 1990; McCafferty *et al.*, 1990; for review see Clackson and Wells, 1994). Similarly, specific peptide ligands have been selected for binding to receptors by affinity selection using large libraries of peptides linked to the C terminus of the lac repressor LacI (Cull *et al.*, 1992). When expressed in *E. coli* the repressor protein physically links the ligand to the encoding plasmid by binding to a lac operator sequence on the plasmid. Moreover, an entirely *in vitro* polysome display system has also been reported (Mattheakis *et al.*, 1994) in which nascent peptides are physically attached via the ribosome to the RNA which encodes them.

In vivo the primary repertoire of antibody specificities is created by a process of DNA rearrangement involving the joining of immunoglobulin V, D and J gene segments. Following antigen encounter in mouse and man, the rearranged V genes in those B cells that have been triggered by the antigen are subjected to a second wave of diversification, this time by somatic hypermutation. This hypermutation generates the secondary repertoire from which good binding specificities can be selected thereby allowing affinity maturation of the humoral immune response.

Artificial selection systems to date rely heavily on initial mutation and selection, similar in concept to the initial phase of V-D-J rearrangement which occurs in natural antibody production, in that it results in the generation of a "fixed" repertoire of gene product mutants from which gene products having the desired activity may be selected.

In vitro RNA selection and evolution (Ellington and Szostak, 1990), sometimes referred to as SELEX (systematic evolution of ligands by exponential enrichment) (Tuerk and Gold, 1990) allows for selection for both binding and chemical activity, but only for nucleic acids. When selection is for binding, a pool of nucleic acids is incubated with immobilised substrate. Non-binders are washed away, then the binders are released, amplified and the whole process is repeated in iterative steps to enrich for better binding sequences. This method can also be adapted to allow isolation of catalytic RNA and DNA (Green and Szostak, 1992; for reviews see Chapman and Szostak, 1994; Joyce, 1994; Gold *et al.*, 1995; Moore, 1995). SELEX, thus, permits cyclical steps of improvement of the desired activity, but is limited in its scope to the preparation of nucleic acids.

Unlike in the natural immune system, however, artificial selection systems are poorly suited to any facile form of "affinity maturation", or cyclical steps of repertoire generation and development. One of the reasons for this is that it is difficult to target mutations to regions of the molecule where they are required, so subsequent cycles of mutation and selection do not lead to the isolation of molecules with improved activity with sufficient efficiency.

Much of what is known about the somatic hypermutation process which occurs during affinity maturation in natural antibody production has been derived from an analysis of the mutations that have occurred during hypermutation *in vivo* (for reviews see Neuberger and Milstein, 1995; Weill and Reynaud, 1996; Parham, 1998). Most of these mutations are single nucleotide substitutions which are introduced in a stepwise manner. They are scattered over the rearranged V domain, though with characteristic hotspots, and the substitutions exhibit a bias for base transitions. The mutations largely accumulate during B cell expansion in germinal centres (rather than during other stages of B cell differentiation and proliferation) with the rate of incorporation of nucleotide substitutions into the V gene during the hypermutation phase estimated at between 10^{-4} and 10^{-3} bp^{-1} $generation^{-1}$ (McKean et al., 1984; Berek & Milstein, 1988)

The possibility that lymphoid cell lines could provide a tractable system for investigating hypermutation was considered many years ago (Coffino and Scharff, 1971; Adetugbo et

al., 1977; Brüggemann et al., 1982). Clearly, it is important that the rate of V gene mutation in the cell-line under study is sufficiently high not only to provide a workable assay but also to be confident that mutations are truly generated by the localised antibody hypermutation mechanism rather than reflecting a generally increased mutation rate as is characteristically associated with many tumours. Extensive studies on mutation have been performed monitoring the reversion of stop codons in V_H in mouse pre-B and plasmacytoma cell lines (Wabl et al., 1985; Chui et al., 1995; Zhu et al., 1995; reviewed by Green et al., 1998). The alternative strategy of direct sequencing of the expressed V gene has indicated that V_H gene diversification in several follicular, Burkitt and Hodgkin lymphomas can continue following the initial transformation event (Bahler and Levy, 1992; Jain et al., 1994; Chapman et al., 1995 and 1996; Braeuninger et al., 1997). Direct sequencing has also revealed a low prevalence of mutations in a cloned follicular lymphoma line arguing that V_H diversification can continue in vitro (Wu et al., 1995). None of the reports of constitutive mutation in cell lines cited above provides evidence that the mutations seen are the result of directed hypermutation, as observed in natural antibody diversification, which is concentrated in the V genes, as opposed to a general susceptibility to mutation as described in many tumour cell lines from different lineages.

Recently, hypermutation has been induced in a cell line by Denepoux *et al.* (1997), by culturing cells in the presence of anti-immunoglobulin antibody and activated T-cells. However, the hypermutation observed was stated to be induced, not constitutive.

Summary of the Invention

In a first aspect of the invention there is provided a method for preparing a lymphoid cell line capable of directed constitutive hypermutation of a target nucleic acid region, comprising screening a cell population for ongoing target sequence diversification, and selecting a cell in which the rate of target nucleic acid mutation exceeds that of other nucleic acid mutation by a factor of 100 or more.

As used herein, "directed constitutive hypermutation" refers to the ability, observed for the first time in experiments reported herein, of certain cell lines to cause alteration of the

nucleic acid sequence of one or more specific sections of endogenous or transgene DNA in a constitutive manner, that is without the requirement for external stimulation. In cells capable of directed constitutive hypermutation, sequences outside of the specific sections of endogenous or transgene DNA are not subjected to mutation rates above background mutation rates.

A "target nucleic acid region" is a nucleic acid sequence or region in the cell according to the invention which is subjected to directed constitutive hypermutation. The target nucleic acid may comprise one or more transcription units encoding gene products, which may be homologous or heterologous to the cell. Exemplary target nucleic acid regions are immunoglobulin V genes as found in immunoglobulin-producing cells These genes are under the influence of hypermutation-recruiting elements, as described further below, which direct the hypermutation to the locus in question. Other target nucleic acid sequences may be constructed, for example by replacing V gene transcription units in loci which contain hypermutation-recruiting elements with another desired transcription unit, or by constructing artificial genes comprising hypermutation-recruiting elements.

"Hypermutation" refers to the mutation of a nucleic acid in a cell at a rate above background. Preferably, hypermutation refers to a rate of mutation of between 10^{-5} and 10^{-3} bp^{-1} $generation^{-1}$. This is greatly in excess of background mutation rates, which are of the order of 10^{-9} to 10^{-10} mutations bp^{-1} $generation^{-1}$ (Drake *et al.*, 1988) and of spontaneous mutations observed in PCR. 30 cycles of amplification with Pfu polymerase would produce $<0.05 \times 10^{-3}$ mutations bp^{-1} in the product, which in the present case would account for less than 1 in 100 of the observed mutations (Lundberg *et al.*, 1991).

Hypermutation is a part of the natural generation of immunoglobulin variable chain (V) genes. According to the present invention therefore, the cell line is preferably an immunoglobulin-producing cell line which is capable of producing at least one immunoglobulin V gene. A V gene may be a variable light chain (V_L) or variable heavy chain (V_H) gene, and may be produced as part of an entire immunoglobulin molecule; it may be a V gene from an antibody, a T-cell receptor or another member of the immunoglobulin superfamily. Members of the immunoglobulin superfamily are involved

in many aspects of cellular and non-cellular interactions *in vivo*, including widespread roles in the immune system (for example, antibodies, T-cell receptor molecules and the like), involvement in cell adhesion (for example the ICAM molecules) and intracellular signalling (for example, receptor molecules, such as the PDGF receptor). Thus, preferred cell lines according to the invention are derived from B-cells. According to the present invention, it has been determined that cell lines derived from antibody-producing B cells may be isolated which retain the ability to hypermutate V region genes, yet do not hypermutate other genes.

In a preferred embodiment, the cells according to the invention are derived from or related to cells which hypermutate *in vivo*. Cells which hypermutate *in vivo* are, for example, immunoglobulin-expressing cells, such as B-cells. Lymphoma cells, which are Ig-expressing cell tumours, are particularly good candidates for the isolation of constitutively hypermutating cell lines according to the present invention.

As used herein, "screening for ongoing target sequence diversification" refers to the determination of the presence of hypermutation in the target nucleic acid region of the cell lines being tested. This can be performed in a variety of ways, including direct sequencing or indirect methods such as the MutS assay (Jolly *et al.*, 1997) or monitoring the generation of immunoglobulin loss variants. Cells selected according to this procedure are cells which display target sequence diversification.

The cell population which is subjected to selection by the method of the invention may be a polyclonal population, comprising a variety of cell types and/or a variety of target sequences, or a (mono-) clonal population of cells.

A clonal cell population is a population of cells derived from a single clone, such that the cells would be identical save for mutations occurring therein. Use of a clonal cell population preferably excludes co-culturing with other cell types, such as activated T-cells, with the aim of inducing V gene hypermutation.

Cells according to the invention do not rely on the use of induction steps in order to produce hypermutation.

Preferably, the clonal cell population screened in the present invention is derived from a B cell. Advantageously it is a lymphoma cell line, such as a Burkitt lymphoma cell line, a follicular lymphoma cell line or a diffuse large cell lymphoma cell line.

Preferably, the method according to the invention further comprises the steps of isolating one or more cells which display target sequence diversification, and comparing the rate of accumulation of mutations in the target sequences with that in non-target sequences in the isolated cells.

A feature of the present invention is that the hypermutation is directed only to specific (target) nucleic acid regions, and is not observed outside of these regions in a general manner. Specificity is thus assayed as part of the method of the invention by assaying the rate of mutation of sequences other than target sequences. C region genes, which are not naturally exposed to hypermutation, may advantageously be employed in such a technique, although any other nucleic acid region not subject to specific hypermutation may also be used. Since hypermutation is not sequence dependent, the actual sequence of the nucleic acid region selected for comparison purposes is not important. However, it must not be subject to control sequences which direct hypermutation, as described below. Conveniently, background mutation may be assessed by fluctuation analysis, for example at the HPRT locus [see Luria and Delbreck, (1943); Capizzi and Jameson, (1973)].

Cells in which target region mutation exceeds non-target region mutation are cells capable of directed constitutive hypermutation of a specific nucleic acid region in accordance with the present invention. The factor by which V region gene mutation exceeds other gene mutation is variable, but is in general of the order of at least 10^2, advantageously 10^3, and preferably 10^4 or more.

Overall mutation rates and diversity may be increased, for example by the administration of mutagens or expression of sequence modifying genes, such as terminal

deoxynucleotidyl transferase (TdT). However, the difference between hypermutation and background is not expected to be increased in such a manner.

Preferred cells according to the invention may be subject to gene manipulation, such as gene deletion, conversion or insertion, in order to increase the rate of somatic hypermutation observed therein. For example, the cells according to the invention may lack one or more copies of a RAD51 paralogue.

The cells may be any suitable vertebrate cells, including mammalian and avian cells.

In a second aspect of the present invention, there is provided a method for preparing a gene product having a desired activity, comprising the steps of:

a) expressing a nucleic acid encoding the gene product in a population of cells according to the first aspect of the present invention, operably linked to a nucleic acid which directs hypermutation;

b) identifying a cell or cells within the population of cells which expresses a mutant gene product having the desired activity; and

c) establishing one or more clonal populations of cells from the cell or cells identified in step (b), and selecting from said clonal populations a cell or cells which expresses a gene product having an improved desired activity.

The population of cells according to part a) above is derived from a clonal or polyclonal population of cells which comprises cells identified by a method according to the first aspect of the invention as being capable of constitutive hypermutation of V region genes. The gene product may thus be the endogenous immunoglobulin polypeptide, a gene product expressed by a manipulated endogenous gene or a gene product expressed by a heterologous transcription unit operatively linked to control sequences which direct somatic hypermutation, as described further below.

The nucleic acid which is expressed in the cells of the invention and subjected to hypermutation may be an endogenous region, such as the endogenous V region, or a heterologous region inserted into the cell line of the invention. This may take form, for

example, of a replacement of the endogenous V region with heterologous transcription unit(s), such as a heterologous V region, retaining the endogenous control sequences which direct hypermutation; or of the insertion into the cell of a heterologous transcription unit under the control of its own control sequences to direct hypermutation, wherein the transcription unit may encode V region genes or any other desired gene product. The nucleic acid according to the invention is described in more detail below.

In step b) above, the cells are screened for the desired gene product activity. This may be, for example in the case of immunoglobulins, a binding activity. Other activities may also be assessed, such as enzymatic activities or the like, using appropriate assay procedures. Where the gene product is displayed on the surface of the cell, cells which produce the desired activity may be isolated by detection of the activity on the cell surface, for example by fluorescence, or by immobilising the cell to a substrate via the surface gene product. Where the activity is secreted into the growth medium, or otherwise assessable only for the entire cell culture as opposed to in each individual cell, it is advantageous to establish a plurality of clonal populations from step a) in order to increase the probability of identifying a cell which secretes a gene product having the desired activity. Advantageously, the selection system employed does not affect the cell's ability to proliferate and mutate.

Preferably, at this stage (and in step c) cells which express gene products having a better, improved or more desirable activity are selected. Such an activity is, for example, a higher affinity binding for a given ligand, or a more effective enzymatic activity. Thus, the method allows for selection of cells on the basis of a qualitative and/or quantitative assessment of the desired activity.

In a third aspect of the present invention, there is provided the use of a cell capable of directed constitutive hypermutation of a specific nucleic acid region in the preparation of a gene product having a desired activity.

In the use according to the invention, a nucleic acid encoding the gene product having the desired activity is operatively linked to control sequences which direct hypermutation

within the cell. Successive generations of the cell thus produce mutants of the nucleic acid sequence, which are screened by the method of the invention to isolate mutants with advantageous properties.

In a further aspect, the invention relates to a cell capable of directed constitutive hypermutation in accordance with the invention. Preferably, the cell is a genetically manipulated chicken DT40 cell. As described above, one or more DNA-repair genes may be manipulated. Preferably, one or more Rad51 genes are manipulated. Advantageously, the genes are downregulated or deleted. Preferably, the genes are Rad51b or Rad51c genes.

In a highly preferred embodiment, the invention provides a cell selected from the group consisting of Δ xrcc2 DT40and Δ xrcc3 DT40

Brief Description of the Figures

Figure 1 V_H diversity in Burkitt lines.

(A) Sequence diversity in the rearranged V_H genes of four sporadic Burkitt lymphoma lines, shown as pie charts. The number of M13 clones sequenced for each cell line is denoted in the centre of the pie; the sizes of the various segments depict the proportion of sequences that are distinguished by 0, 1, 2 etc. mutations (as indicated) from the consensus.

(B) Presumed dynastic relationship of V_H mutations identified in the initial Ramos culture. Each circle (with shading proportional to extent of mutation) represents a distinct sequence with the number of mutations accumulated indicated within the circle.

(C) Mutation prevalence in the rearranged V_λ genes. Two V_λ rearrangements are identified in Ramos. Diversity and assignment of germline origin is presented as in Figure 1A.

(D) Comparison of mutation prevalence in the V_H and $C\mu$ regions of the initial Ramos culture. Pie charts are presented as in Figure 1A.

Figure 2 Constitutive V_H diversification in Ramos.

(A) Diversification assessed by a MutS assay: The mutation prevalence in each population as deduced by direct cloning and sequencing is indicated.

(B) Dynastic relationships deduced from the progeny of three independent Ramos clones.

Figure 3. Distribution of unselected nucleotide substitutions along the Ramos V_H.

Figure 4. Hypermutation in Ramos generates diverse revertible IgM-loss variants.

(A) Scheme showing the isolation of IgM-loss variants.

(B) Table showing that multiple nonsense mutations can contribute to V_H inactivation. Each V_H codon position at which stops are observed in these two populations is listed.

(C) Table of reversion rates of IgM-loss variants.

(D) Sequence surrounding the stop codons in the IgM-loss derivatives.

Figure 5. IgM-loss variants in Ramos transfectants expressing TdT.

(A) Western blot analysis of expression of TdT in three pSV-pβG/TdT and three control transfectants of Ramos.

(B) Pie charts depicting independent mutational events giving rise to IgM-loss variants.

Figure 6. Sequence table summarising mutations in V_H other than single nucleotide substitutions.

Figure 7. Comparison of sequences isolated from V_H genes of Ramos cells which have lost anti-idiotype (anti-Id1) binding specificity. Nucleotide substitutions which differ from the starting population consensus are shown in bold. Predicted amino acid changes are indicated, also in bold type.

Figure 8. Bar graph showing enrichment of Ramos cells for production of an immunoglobulin with a novel binding specificity, by iterative selection over five rounds.

Figure 9. Bar graph showing improved recovery of Ramos cells binding a novel specificity (streptavidin) by increasing the bead:cell ratio.

Figure 10. Chart showing increase in recovery of novel binding specificity Ramos cells according to increasing target antigen concentration.

Figure 11. V_H sequence derived from streptavidin-binding Ramos cells. Nucleotide changes observed in comparison with the V_H sequence of the starting population, and predicted amino acid changes, are shown in bold.

Figure 12. Amount of IgM in supernatants of cells selected in rounds 4, 6 and 7 of a selection process for streptavidin binding, against control medium and unselected Ramos cell supernatant.

Figure 13. Streptavidin binding of IgM from the supernatants of Figure 12.

Figure 14. Streptavidin binding of supernatants from round 4 and round 6 of a selection for streptavidin binding, analysed by surface plasmon resonance.

Figure 15. FACS analysis of binding to streptavidin-FITC of cells selected in rounds 4 and 6.

Figure 16. V_H and V_L sequences of round 6 selected IgM.

Figure 17. FACS analysis of affinity matured Ramos cells selected against streptavidin.

Figure 18. ELISA of affinity matured Ramos cells.

Figure 19. sIgM-loss variants in wild-type and repair deficient DT40.

(A) flow cytometric analysis of sIgM heterogeneity in wild type and repair deficient cells.

(B) fluctuation analysis of the frequency of generation of sIgM-loss variants.

Figure 20. Analysis of V_λ sequences cloned from sIgM variants of DT40.

Figure 21. Analysis of Ig sequences of unsorted DT40 populations after one month of clonal expansion.

Figure 22. Analysis of sIgM loss variants of DT40 cells deficient in DNA-PK, Ku70 and Rad51B.

Figure 23. A dynasty of IgMs from DT40 specific for a rat immunoglobulin idiotype. Cells were subjected to six rounds of selection using an aggregate of biotinylated-rat S7 IgG2a, | monoclonal antibody (Ab)/FITC-streptavidin to yield DT-Ab3, which was then subjected to a further three rounds of sorting with a direct PE-Ab conjugate to yield DTAb6. **(a)** Analysis of the binding of biotinylated Ab/FITC-Strep aggregate as well as of PE-Ab to DT-Ab3 and DT-Ab6 cells. **(b)** Binding of IgM in the supernatant of DTAb3,5 and 6 cells to the S7 Ab monitored by ELISA using plates coated with S7 Ab and detection with anti-chicken IgM. Supernatants were controlled for total IgM titres. No binding by the DT-Ab6 IgM was detected to a wide variety of rat hybridoma and chimaeric mAbs of different isotypes (generously provided by G. Butcher and M Brüggemann (not shown)). **(c)** Binding of DT-Ab6 IgM to S7 rat IgG Ab coated on to the plate is competed by S7 Ab itself but scarcely by normal rat serum. The S7 (stock at 0.5 mg/ml) and rat serum competitors were used at the dilutions indicated. **(d)** DT-Ab6 IgM can be used to stain S7 hybridoma cells. Staining was detected by flow cytometry of fixed, permeabilised cells, detecting using FITC-conjugated anti-chicken IgM. **(e)** Comparison of the staining of DT-Ab3 and DT-Ab6 cells by PE-Ab conjugate using various dilutions of the original 0.2 mg/ml stock (PE-S7 Ab; Pharamacia). **(f)** Affinity determination. DT-Ab6 cells (10^6) were incubated with 0.2 picomoles of PE-conjugated S7 Ab in various volumes to give the PE-Ab concentrations indicated; the mean

fluorescence intensity (MFI) of the cells determined following washing. (g) Comparison of the IgV$_H$/V$_L$ sequences of DT-Ab6 and the parental population.

Figure 24. A dynasty of Protein A-specific IgMs from DT40. (a) Selection of DT40 variants binding to derivatised magnetic beads showing the number of cells recovered following incubation of 10_8 cells with 10_6 beads. DT-P1 and P2 were selected on streptavidin-beads coated with biotinylated-Protein A whereas DT-P3 was selected from DT-P2 on tosylated-magnetic beads directly coated with Protein A. (b) Comparison of staining of subclones of various DT-P cells by an aggregate of biotinylated Protein A with FITC-streptavidin, by Protein A that had been directly conjugated with FITC and by FITC-anti-IgM. (c) Binding of IgM in the supernatant of parental DT40 and DT-P14 bulk population to Protein A-coated plates monitored by ELISA using a mouse IgM mAb specific for chicken IgM (Southern Biotechnology) for detection. Total IgM titres were also compared as a control. (d) IgM from the supernatant of DT-P4, -P6 and P7 clones (but not from parental DT40) can be purified on Protein A, using Western blot to detect chicken IgM retained on the Protein A-Sepharose. (e) Binding of [^{35}S]Protein A to DTP cells. Parental DT40 (DT) or DT-P cells ($5x10_5$ cells in 0.1 ml)) were incubated on ice for 1h with [^{35}S]Protein A (48,000 bindable cpm; 200-2000 Ci/mmol); bound cpm were determined following a single PBS wash. (f) Affinity determination. DT-P14 cells (10_6) were incubated with 0.7 picomoles [^{35}S]Protein A ($7x10_{10}$ bindable cpm/μmole) in various volumes to yield the Protein A concentrations indicated; bound cpm were determined following washing. (g) Staining of DT-P4 and DT-P14 cells with FITCProtein A is enhanced by performing the staining in the presence (P4+, P14+) of 1mg/ml unlabelled rabbit IgG. (h) IgV gene sequences from DT-P subclones.

Figure 25. Antigen-selection of DT40 can yield polyreactive IgMs with the possibility of maturation into specificity. DT-H10 cells were obtained by ten sequential rounds of selection from the parental Xrcc2-deficient DT40 pool using carboxylated magnetic beads to which human serum albumin (HSA) had been coupled. DT-T10 and DT-O6 cells were obtained by serial rounds of flow cytometric selection of cells stained with aggregates of FITC-streptavidin with either biotinylated thyrogobulin (DT-T cells) or ovalbumin (DT-O cells), taking the brightest 2% of the population in each round. (a) Flow cytometric

analysis of DT-H6 and DT-T6 cells stained with the reagents indicated (Cy-strep, cychrome-streptavidin). (b) Antigen binding by DT-H6 and DT-T6 cells is mediated by the surface IgM as witnessed by 2D flow cytometric analysis of DT-H6 and DT-T6 subpopulations that have been enriched for sIgM-loss variants. (c) ELISA of DTH6 and DT-T6 culture supernatants, monitoring chicken IgM specific for the antigens indicated. (d) DT-O6 cells (selected with ovalbumin aggregates) also bind Tg (as well as several other antigens tested (not shown)) but sequential sorting for FITC-Ova_{bright} /$Cystreptavidin_{dull}$ cells yields a population exhibiting greater specificity for Ova.

Figure 26. Analysis of naturally-occuring constitutively hypermutating BL cell lines.

Detailed Description of the Invention

The present invention makes available for the first time a cell line which constitutively hypermutates selected nucleic acid regions. This permits the design of systems which produce mutated gene products by a technique which mirrors affinity maturation in natural antibody production. The Ramos Burkitt line constitutively diversifies its rearranged immunoglobulin V gene during in vitro culture. This hypermutation does not require stimulation by activated T cells, exogenously-added cytokines or even maintenance of the B cell antigen receptor.

The rate of mutation (which lies in the range $0.2\text{-}1\times10^{-4}$ bp^{-1} $generation^{-1}$) is sufficiently high to readily allow the accumulation of a large database of unselected mutations and so reveal that hypermutation in Ramos exhibits most of the features classically associated with immunoglobulin V gene hypermutation in vivo (preferential targeting of mutation to the V; stepwise accumulation of single nucleotide substitutions; transition bias; characteristic mutational hotspots). The large majority of mutations in the unselected database are single nucleotide substitutions although deletions and duplications (sometimes with a flanking nucleotide substitution) are detectable. Such deletions and duplications have also been proposed to be generated as a consequence of hypermutation in vivo (Wilson et al., 1998; Goosens et al., 1998; Wu & Kaartinen, 1995).

The isolation of cells which constitutively hypermutate selected nucleic acid regions is based on the monitoring of V gene mutation in cell lines derived from antibody-producing cells such as B cells. The selection method employed in the invention may be configured in a number of ways.

Selection of Hypermutating Cells

Hypermutating cells may be selected from a population of cells by a variety of techniques, including sequencing of target sequences, selection for expression loss mutants, assay using bacterial MutS protein and selection for change in gene product activity.

One of the features of hypermutation of target nucleic acids is that the process results in the introduction of stop codons into the target sequence with far greater frequency than would be observed in the absence of hypermutation. This results in loss of production of a gene product from the cell. This loss may be exploited to identify cells which are hypermutating nucleic acid sequences.

In a preferred embodiment of the invention, the target nucleic acid encodes an immunoglobulin. Immunoglobulin loss may be detected both for cells which secrete immunoglobulins into the culture medium, and for cells in which the immunoglobulin is displayed on the cell surface. Where the immunoglobulin is present on the cell surface, its absence may be identified for individual cells, for example by FACS analysis, immunofluorescence microscopy or ligand immobilisation to a support. In a preferred embodiment, cells may be mixed with antigen-coated magnetic beads which, when sedimented, will remove from the cell suspension all cells having an immunoglobulin of the desired specificity displayed on the surface.

The technique may be extended to any immunoglobulin molecule, including antibodies, T-cell receptors and the like. The selection of immunoglobulin molecules will depend on the nature of the clonal population of cells which it is desired to assay according to the invention.

Alternatively, cells according to the invention may be selected by sequencing of target nucleic acids, such as V genes, and detection of mutations by sequence comparison. This process may be automated in order to increase throughput.

In a further embodiment, cells which hypermutate V genes may be detected by assessing change in antigen binding activity in the immunoglobulins produced in a clonal cell population. For example, the quantity of antigen bound by a specific unit amount of cell medium or extract may be assessed in order to determine the proportion of immunoglobulin produced by the cell which retains a specified binding activity. As the V genes are mutated, so binding activity will be varied and the proportion of produced immunoglobulin which binds a specified antigen will be reduced.

Alternatively, cells may be assessed in a similar manner for the ability to develop a novel binding affinity, such as by exposing them to an antigen or mixture of antigens which are initially not bound and observing whether a binding affinity develops as the result of hypermutation.

In a further embodiment, the bacterial MutS assay may be used to detect sequence variation in target nucleic acids. The MutS protein binds to mismatches in nucleic acid hybrids. By creating heteroduplexes between parental nucleic acids and those of potentially mutated progeny, the extent of mismatch formation, and thus the extent of nucleic acid mutation, can be assessed.

Where the target nucleic acid encodes an gene product other than an immunoglobulin, selection may be performed by screening for loss or alteration of a function other than binding. For example, the loss or alteration of an enzymatic activity may be screened for.

Cells which target sequence hypermutation are assessed for mutation in other nucleic acid regions. A convenient region to assay is the constant (C) region of an immunoglobulin gene. C regions are not subject to directed hypermutation according to the invention. The assessment of C regions is preferably made by sequencing and comparison, since this is

the most certain method for determining the *absence* of mutations. However, other techniques may be employed, such as monitoring for the retention of C region activities, for example complement fixation, which may be disrupted by hypermutation events.

Genetic Manipulation of cells

Hypermutating cells according to the invention may be selected from cells which have been genetically manipulated to enhance rates of hypermutation in the Ig V-region. Genes which are responsible for modulation of mutation rates include, in general, in nucleic acid repair procedures in the cell. Genes which are manipulated in accordance with the present invention may be upregulated, downregulated or deleted.

Up- or down-regulation refers to an increase, or decrease, in activity of the gene product encoded by the gene in question by at least 10%, preferably 25%, more preferably 40, 50, 60, 70, 80, 90,95, 99% or more. Upregulation may of course represent an increase in activity of over 100%, such as 200% or 500%. A gene which is 100% downregulated is functionally deleted and is referred to herein as "deleted".

Preferred genes manipulated in accordance with the present invention include analogues and/or paralogues of the Rad51 gene, in particular xrcc2, xrcc3 and Rad51bgenes.

Rad51 analogues and/or paralogues are advantageously downregulated, and preferably deleted. Downregulation or deletion of one or more Rad51 paralogues, gives rise to an increase in hypermutation rates in accordance with the invention. Preferably, two or more Rad51 genes, including analogues and/or paralogues thereof, are downregulated or deleted.

In a highly preferred embodiment, avian cell lines such as the chicken DT40 cell line are modified by deletion of xrcc2 and/or xrcc3. Δ xrcc2 DT40 as well Δxrcc3-DT40 are constitutively hypermutating cell lines isolated in accordance with the present invention.

Adaptation of the endogenous gene products

Having obtained a cell line which constitutively hypermutates an endogenous gene, such as an immunoglobulin V region gene, the present invention provides for the adaptation of the endogenous gene product, by constitutive hypermutation, to produce a gene product having novel properties. For example, the present invention provides for the production of an immunoglobulin having a novel binding specificity or an altered binding affinity.

The process of hypermutation is employed, in nature, to generate improved or novel binding specificities in immunoglobulin molecules. Thus, by selecting cells according to the invention which produce immunoglobulins capable of binding to the desired antigen and then propagating these cells in order to allow the generation of further mutants, cells which express immunoglobulins having improved binding to the desired antigen may be isolated.

A variety of selection procedures may be applied for the isolation of mutants having a desired specificity. These include Fluorescence Activated Cell Sorting (FACS), cell separation using magnetic particles, antigen chromatography methods and other cell separation techniques such as use of polystyrene beads.

Separating cells using magnetic capture may be accomplished by conjugating the antigen of interest to magnetic particles or beads. For example, the antigen may be conjugated to superparamagnetic iron-dextran particles or beads as supplied by Miltenyi Biotec GmbH. These conjugated particles or beads are then mixed with a cell population which may express a diversity of surface immunoglobulins. If a particular cell expresses an immunoglobulin capable of binding the antigen, it will become complexed with the magnetic beads by virtue of this interaction. A magnetic field is then applied to the suspension which immobilises the magnetic particles, and retains any cells which are associated with them via the covalently linked antigen. Unbound cells which do not become linked to the beads are then washed away, leaving a population of cells which is isolated purely on its ability to bind the antigen of interest. Reagents and kits are available from various sources for performing such one-step isolations, and include Dynal

Beads (Dynal AS; http://www.dynal.no), MACS-Magnetic Cell Sorting (Miltenyi Biotec GmbH; http://www.miltenyibiotec.com), CliniMACS (AmCell; http://www.amcell.com) as well as Biomag, Amerlex-M beads and others.

Fluorescence Activated Cell Sorting (FACS) can be used to isolate cells on the basis of their differing surface molecules, for example surface displayed immunoglobulins. Cells in the sample or population to be sorted are stained with specific fluorescent reagents which bind to the cell surface molecules. These reagents would be the antigen(s) of interest linked (either directly or indirectly) to fluorescent markers such as fluorescein, Texas Red, malachite green, green fluorescent protein (GFP), or any other fluorophore known to those skilled in the art. The cell population is then introduced into the vibrating flow chamber of the FACS machine. The cell stream passing out of the chamber is encased in a sheath of buffer fluid such as PBS (Phosphate Buffered Saline). The stream is illuminated by laser light and each cell is measured for fluorescence, indicating binding of the fluorescent labelled antigen. The vibration in the cell stream causes it to break up into droplets, which carry a small electrical charge. These droplets can be steered by electric deflection plates under computer control to collect different cell populations according to their affinity for the fluorescent labelled antigen. In this manner, cell populations which exhibit different affinities for the antigen(s) of interest can be easily separated from those cells which do not bind the antigen. FACS machines and reagents for use in FACS are widely available from sources world-wide such as Becton-Dickinson, or from service providers such as Arizona Research Laboratories (http://www.arl.arizona.edu/facs/).

Another method which can be used to separate populations of cells according to the affinity of their cell surface protein(s) for a particular antigen is affinity chromatography. In this method, a suitable resin (for example CL-600 Sepharose, Pharmacia Inc.) is covalently linked to the appropriate antigen. This resin is packed into a column, and the mixed population of cells is passed over the column. After a suitable period of incubation (for example 20 minutes), unbound cells are washed away using (for example) PBS buffer. This leaves only that subset of cells expressing immunoglobulins which bound the antigen(s) of interest, and these cells are then eluted from the column using (for example)

an excess of the antigen of interest, or by enzymatically or chemically cleaving the antigen from the resin. This may be done using a specific protease such as factor X, thrombin, or other specific protease known to those skilled in the art to cleave the antigen from the column via an appropriate cleavage site which has previously been incorporated into the antigen-resin complex. Alternatively, a non-specific protease, for example trypsin, may be employed to remove the antigen from the resin, thereby releasing that population of cells which exhibited affinity for the antigen of interest.

Insertion of heterologous transcription units

In order to maximise the chances of quickly selecting an antibody variant capable of binding to any given antigen, or to exploit the hypermutation system for non-immunoglobulin genes, a number of techniques may be employed to engineer cells according to the invention such that their hypermutating abilities may be exploited.

In a first embodiment, transgenes are transfected into a cell according to the invention such that the transgenes become targets for the directed hypermutation events.

As used herein, a "transgene" is a nucleic acid molecule which is inserted into a cell, such as by transfection or transduction. For example, a "transgene" may comprise a heterologous transcription unit as referred to above, which may be inserted into the genome of a cell at a desired location.

The plasmids used for delivering the transgene to the cells are of conventional construction and comprise a coding sequence, encoding the desired gene product, under the control of a promoter. Gene transcription from vectors in cells according to the invention may be controlled by promoters derived from the genomes of viruses such as polyoma virus, adenovirus, fowlpox virus, bovine papilloma virus, avian sarcoma virus, cytomegalovirus (CMV), a retrovirus and Simian Virus 40 (SV40), from heterologous mammalian promoters such as the actin promoter or a very strong promoter, e.g. a ribosomal protein promoter, and from the promoter normally associated with the

heterologous coding sequence, provided such promoters are compatible with the host system of the invention.

Transcription of a heterologous coding sequence by cells according to the invention may be increased by inserting an enhancer sequence into the vector. Enhancers are relatively orientation and position independent. Many enhancer sequences are known from mammalian genes (e.g. elastase and globin). However, typically one will employ an enhancer from a eukaryotic cell virus. Examples include the SV40 enhancer on the late side of the replication origin (bp 100-270) and the CMV early promoter enhancer. The enhancer may be spliced into the vector at a position 5' or 3' to the coding sequence, but is preferably located at a site 5' from the promoter.

Advantageously, a eukaryotic expression vector may comprise a locus control region (LCR). LCRs are capable of directing high-level integration site independent expression of transgenes integrated into host cell chromatin, which is of importance especially where the heterologous coding sequence is to be expressed in the context of a permanently-transfected eukaryotic cell line in which chromosomal integration of the vector has occurred, in vectors designed for gene therapy applications or in transgenic animals.

Eukaryotic expression vectors will also contain sequences necessary for the termination of transcription and for stabilising the mRNA. Such sequences are commonly available from the 5' and 3' untranslated regions of eukaryotic or viral DNAs or cDNAs. These regions contain nucleotide segments transcribed as polyadenylated fragments in the untranslated portion of the mRNA.

An expression vector includes any vector capable of expressing a coding sequence encoding a desired gene product that is operatively linked with regulatory sequences, such as promoter regions, that are capable of expression of such DNAs. Thus, an expression vector refers to a recombinant DNA or RNA construct, such as a plasmid, a phage, recombinant virus or other vector, that upon introduction into an appropriate host cell, results in expression of the cloned DNA. Appropriate expression vectors are well known to those with ordinary skill in the art and include those that are replicable in eukaryotic

and/or prokaryotic cells and those that remain episomal or those which integrate into the host cell genome. For example, DNAs encoding a heterologous coding sequence may be inserted into a vector suitable for expression of cDNAs in mammalian cells, e.g. a CMV enhancer-based vector such as pEVRF (Matthias, et al., 1989).

Construction of vectors according to the invention employs conventional ligation techniques. Isolated plasmids or DNA fragments are cleaved, tailored, and religated in the form desired to generate the plasmids required. If desired, analysis to confirm correct sequences in the constructed plasmids is performed in a known fashion. Suitable methods for constructing expression vectors, preparing in vitro transcripts, introducing DNA into host cells, and performing analyses for assessing gene product expression and function are known to those skilled in the art. Gene presence, amplification and/or expression may be measured in a sample directly, for example, by conventional Southern blotting, Northern blotting to quantitate the transcription of mRNA, dot blotting (DNA or RNA analysis), or in situ hybridisation, using an appropriately labelled probe which may be based on a sequence provided herein. Those skilled in the art will readily envisage how these methods may be modified, if desired.

In one variation of the first embodiment, transgenes according to the invention also comprise sequences which direct hypermutation. Such sequences have been characterised, and include those sequences set forth in Klix *et al.*, (1998), and Sharpe *et al.*, (1991), incorporated herein by reference. Thus, an entire locus capable of expressing a gene product and directing hypermutation to the transcription unit encoding the gene product is transferred into the cells. The transcription unit and the sequences which direct hypermutation are thus exogenous to the cell. However, although exogenous the sequences which direct hypermutation themselves may be similar or identical to the sequences which direct hypermutation naturally found in the cell

In a second embodiment, the endogenous V gene(s) or segments thereof may be replaced with heterologous V gene(s) by homologous recombination, or by gene targeting using, for example, a Lox/Cre system or an analogous technology or by insertion into hypermutating cell lines which have spontaneously deleted endogenous V genes.

Alternatively, V region gene(s) may be replaced by exploiting the observation that hypermutation is accompanied by double stranded breaks in the vicinity of rearranged V genes.

The invention is further described below, for the purposes of illustration only, in the following examples.

Example 1: Selection of a hypermutating cell

In order to screen for a cell that undergoes hypermutation in vitro, the extent of diversity that accumulates in several human Burkitt lymphomas during clonal expansion is assessed. The Burkitt lines BL2, BL41 and BL70 are kindly provided by G. Lenoir (IARC, Lyon, France) and Ramos (Klein et al., 1975) is provided by D. Fearon (Cambridge, UK). Their rearranged V_H genes are PCR amplified from genomic DNA using multiple V_H family primers together with a J_H consensus oligonucleotide. Amplification of rearranged V_H segments is accomplished using Pfu polymerase together with one of 14 primers designed for each of the major human V_H families (Tomlinson, 1997) and a consensus J_H back primer which anneals to all six human J_H segments (JOL48, 5'-GCGGTACCTGAGGAGACGGTGACC-3', gift of C. Jolly). Amplification of the Ramos V_H from genomic DNA is performed with oligonucleotides RVHFOR (5'-CCCCAAGCTTCCCAGGTGCAGCTACAGCAG) and JOL48. Amplification of the expressed V_H-Cμ cDNA is performed using RVHFOR and Cμ2BACK (5'-CCCCGGTACCAGATGAGCTTGGACTTGCGG). The genomic Cμ1/2 region is amplified using Cμ2BACK with Cμ1FOR (5'-CCCCAAGCTTCGGGAGTGCATCCGCCCCAACCCTT); the functional Cμ allele of Ramos contains a C at nucleotide 8 of Cμ2 as opposed to T on the non-functional allele. Rearranged Vλs are amplified using 5'-CCCCAAGCTTCCCAGTCTGCCCTGACTCAG and 5'-CCCCTCTAGACCACCTAGGACGGTC-AGCTT. PCR products are purified using QIAquick (Qiagen) spin columns and sequenced using an ABI377 sequencer following cloning into M13. Mutations are computed using the GAP4 alignment program (Bonfield et al., 1995).

Sequencing of the cloned PCR products reveals considerable diversity in the Ramos cell line (a prevalence of $2.8x10^{-3}$ mutations bp^{-1} in the V_H) although significant heterogeneity is also observed in BL41 as well as in BL2. See Figure 1A. Sequence diversity in the rearranged V_H genes of four sporadic Burkitt lymphoma lines are shown as pie charts. The rearranged V_H genes in each cell line are PCR amplified and cloned into M13. For each cell line, the consensus is taken as the sequence common to the greatest number of M13 clones and a germline counterpart (indicated above each pie) assigned on the basis of closest match using the VBASE database of human immunoglobulin sequences (Tomlinson, 1997). The V_H consensus sequence for Ramos used herein differs in 3 positions from the sequence determined by Chapman et al (1996), five positions from that determined by Ratech (1992) and six positions from its closest germline counterpart V_H4(DP-63).

The analysis of V_H diversity in Ramos is extended by sequencing the products from nine independent PCR amplifications. This enables a likely dynastic relationship between the mutated clones in the population to be deduced, minimising the number of presumed independent repeats of individual nucleotide substitutions (Figure 1B). 315 M13V_H clones obtained from nine independent PCR amplifications are sequenced; the dynasty only includes sequences identified (rather than presumed intermediates). Individual mutations are designated according to the format "C230" with 230 being the nucleotide position in the Ramos V_H (numbered as in Figure 3) and the "C" indicating the novel base at that position. The criterion used to deduce the genealogy is a minimisation of the number of independent occurrences of the same nucleotide substitution. The majority of branches contain individual members contributed by distinct PCR amplifications. The rare deletions and duplications are indicated by the prefix "x" and "d" respectively. Arrows highlight two mutations (a substitution at position 264 yielding a stop codon and a duplication at position 184) whose position within the tree implies that mutations can continue to accumulate following loss of functional heavy chain expression.

PCR artefacts make little contribution to the database of mutations; not only is the prevalence of nucleotide substitutions greatly in excess of that observed in control PCR amplifications ($<0.05 \times 10^{-3}$ bp^{-1}) but also identically mutated clones (as well as dynastically related ones) are found in independent amplifications. In many cases, generations within a lineage differ by a single nucleotide substitution indicating that only a small number of substitutions have been introduced in each round of mutation.

Analysis of Vλ rearrangements reveals that Ramos harbours an in-frame rearrangement of Vλ2.2-16 (as described by Chapman et al. 1996)) and an out-of-frame rearrangement of Vλ2.2-25. There is mutational diversity in both rearranged Vλs although greater diversity has accumulated on the non-functional allele (Figure 1C).

A classic feature of antibody hypermutation is that mutations largely accumulate in the V region but scarcely in the C. This is also evident in the mutations that have accumulated in the Ramos IgH locus (Figure 1D). M13 clones containing cDNA inserts extending through V_H, Cμ1 and the first 87 nucleotides Cμ2 are generated by PCR from the initial Ramos culture. The Pie charts (presented as in Figure 1A) depict the extent of mutation identified in the 341 nucleotide stretch of V_H as compared to a 380 nucleotide stretch of Cμ extending from the beginning of Cμ1.

The IgM immunoglobulin produced by Ramos is present both on the surface of the cells and, in secreted form, in the culture medium. Analysis of the culture medium reveals that Ramos secretes immunoglobulin molecules to a very high concentration, approximately 1μg/ml. Thus, Ramos is capable of secreting immunoglobulins to a level which renders it unnecessary to reclone immunoglobulin genes into expression cell lines or bacteria for production.

Example 2: V_H diversification in Ramos is constitutive

To address whether V gene diversification is ongoing, the cells are cloned and V_H diversity assessed using a MutS-based assay after periods of *in vitro* culture. The Ramos V_H is PCR amplified and purified as described above using oligonucleotides containing a biotinylated base at the 5'-end. Following denaturation/renaturation (99^{O}C for 3 min; 75 ^{O}C for 90 min), the extent of mutation is assessed by monitoring the binding of the mismatched heteroduplexed material to the bacterial mismatch-repair protein MutS, filter-bound, with detection by ECL as previously described (Jolly et al., 1997).

The results indicate that V_H diversification is indeed ongoing (see Figure 2A). DNA is extracted from Ramos cells that have been cultured for 1 or 3 months following limit dilution cloning. The rearranged V_H is PCR amplified using biotinylated oligonucleotides prior to undergoing denaturation/renaturation; mismatched heteroduplexes are then detected by binding to immobilised MutS as previously described (Jolly et al., 1997). An aliquot of the renatured DNA is bound directly onto membranes to confirm matched DNA loading (Total DNA control). Assays performed on the Ramos V_H amplified from a bacterial plasmid template as well as from the initial Ramos culture are included for comparison.

The V_H genes are PCR amplified from Ramos cultures that have been expanded for four (Rc1) or six (Rc13 and 14) weeks (Figure 2B). A mutation rate for each clone is indicated and is calculated by dividing the prevalence of independent V_H mutations at 4 or 6 weeks post-cloning by the presumed number of cell divisions based on a generation time of 24 h. The sequences reveal step-wise mutation accumulation with a mutation rate of about 0.24×10^{-4} mutations bp^{-1} $generation^{-1}$.

Direct comparison of the V_H mutation rate in Ramos to that in other cell-lines is not straightforward since there is little information on mutation rates in other lines as judged by unselected mutations incorporated throughout the V_H obtained following clonal expansion from a single precursor cell. However, the prevalence of mutations following a

two week expansion of 50 precursor BL2 cells has been determined under conditions of mutation induction ($2.7x10^{-3}$ mutations bp^{-1}; Denépoux et al., 1997). Similar experiments performed with Ramos under conditions of normal culture reveal a mutation prevalence of $2.3x10^{-3}$ mutations bp^{-1}. Various attempts to enhance the mutation rate by provision of cytokines, helper T cells etc. have proved unsuccessful. Thus, the rate of mutation that can be achieved by specific induction in BL2 cells appears to be similar to the constitutive rate of V_H mutation in Ramos.

Example 3: Examination of the nature of V_H mutations in Ramos

A database of mutational events is created which combines those detected in the initial Ramos culture (from 141 distinct sequences) with those detected in four subclones that have been cultured in various experiments without specific selection (from a further 135 distinct sequences). This database is created after the individual sets of sequences have been assembled into dynastic relationships (as detailed in the legend to Figure 1B) to ensure that clonal expansion of an individual mutated cell does not lead to a specific mutational event being counted multiple times. Here an analysis of this composite database of 340 distinct and presumably unselected mutational events (200 contributed by the initial Ramos culture and 140 from the expanded subclones) is described; separate analysis of the initial and subclone populations yields identical conclusions.

The overwhelming majority of the mutations (333 out of 340) are single nucleotide substitutions. A small number of deletions (4) and duplications (3) are observed but no untemplated insertions; these events are further discussed below. There are only five sequences which exhibited nucleotide substitutions in adjacent positions; however, in three of these five cases, the genealogy revealed that the adjacent substitutions have been sequentially incorporated. Thus, the simultaneous creation of nucleotide substitutions in adjacent positions is a rare event.

The distribution of the mutations along the V_H is highly non-random (See Figure 3). Independently occurring base substitutions are indicated at each nucleotide position. The locations of CDR1 and 2 are indicated. Nucleotide positions are numbered from the 3'-

end of the sequencing primer with nucleotide position +1 corresponding to the first base of codon 7; codons are numbered according to Kabat. Mutations indicated in italics (nucleotide position 15, 193, 195 and 237) are substitutions that occur in a mutated subclone and have reverted the sequence at that position to the indicated consensus.

The major hotspot is at the G and C nucleotides of the Ser82a codon, which has previously been identified as a major intrinsic mutational hotspot in other V_H genes (Wagner et al., 1995; Jolly et al., 1996) and conforms to the RGYW consensus (Rogozin and Kolchanov, 1992; Betz et al., 1993). Whilst the dominant intrinsic mutational hotspot in many V_H genes is at Ser31, this codon is not present in the Ramos consensus V_H (or its germline counterpart) which have Gly at that position. The individual nucleotide substitutions show a marked bias in favour of transitions (51% rather than randomly-expected 33%). There is also a striking preference for targeting G and C which account for 82% of the nucleotides targeted (Table 1).

Table 1. Nucleotide substitution preferences of hypermutation in Ramos

	Frequency of substitution to				
Parental nucleotide	T	C	G	A	Total
T	-	3.9	1.2	3.0	8.1
C	17.4	-	12.6	4.8	34.8
G	7.2	15.9	-	24.0	47.1
A	2.4	1.8	5.7	-	9.9

Single nucleotide substitutions were computed on the V_H coding strand and are given as the percentage of the total number (333) of independent, unselected nucleotide substitutions identified.

Example 4: Selection of hypermutating cells by IgM-loss

Analysis of the Ramos variants reveals several mutations that must have inactivated V_H (see Figure 1B) suggesting it might be possible for the cells to lose IgM expression but remain viable. If this is the case, Ig expression loss would be an easy means to select a constitutively hypermutating B cell line.

Analysis of the Ramos culture reveals it to contain 8% surface IgM^- cells. Such IgM-loss variants are generated during in vitro culture, as follows. The starting Ramos culture is transfected with a pSV2neo plasmid, diluted into 96-well plates and clones growing in selective medium allowed to expand. Flow cytometry performed on the expanded clones six months after the original transfection reveals the presence of IgM-loss variants,

constituting 16% and 18% of the two clonal populations (Rc13 and Rc14) shown here (Figure 4A). Enrichment by a single round of sorting yields subpopulations that contain 87% (Rc13) and 76% (Rc14) surface IgM-negative cells. Following PCR amplification of the rearranged V_H gene in these subpopulations, sequencing reveals that 75% (Rc13) and 67% (Rc14) of the cloned V_H segments contained a nonsense (stop), deletion (del) or duplication (dup) mutation within the 341 nucleotide V_H stretch analysed. The remainder of the clones are designated wild type (wt) although no attempt is made to discriminate possible V_H-inactivating missense mutations. The 4 deletions and 3 duplications identified in the Rc13 population are all distinct whereas only 4 distinct mutations account for the 7 Rc14 sequences determined that harbour deletions. The nature of the deletions and duplications is presented in Figure 6: each event is named with a letter followed by a number. The letter gives the provenance of the mutation (A, B and C being the cloned TdT^- control transfectants, D, E and F the TdT^+ transfectants and U signifies events identified in the initial, unselected Ramos culture); the number indicates the first nucleotide position in the sequence string. Nucleotides deleted are specified above the line and nucleotides added (duplications or non-templated insertions) below the line; single nucleotide substitutions are encircled with the novel base being specified. The duplicated segments of V_H origin are underlined; non-templated insertions are in bold. With several deletions or duplications, the event is flanked by a single nucleotide of unknown provenance. Such flanking changes could well arise by nucleotide substitution (rather than non-templated insertion) and these events therefore separately grouped; the assignment of the single base substitution (encircled) to one or other end of the deletion/duplication is often arbitrary.

The IgM^- cells are enriched in a single round of sorting prior to PCR amplification and cloning of their V_H segments. The sequences reveal a considerable range of V_H-inactivating mutations (stop codons or frameshifts) (Figure 4) although diverse inactivating mutations are even evident in IgM-loss variants sorted after only 6 weeks of clonal expansion (see Figure 5). In Figure 5A expression of TdT in three pSV-pβG/TdT and three control transfectants of Ramos is compared by Western blot analysis of nuclear

protein extracts. Nalm6 (a TdT-positive human pre-B cell lymphoma) and HMy2 (a TdT-negative mature human B lymphoma) provided controls.

In Figure 5B, pie charts are shown depicting independent mutational events giving rise to IgM-loss variants. IgM^- variants (constituting 1-5% of the population) are obtained by sorting the three TdT^+ and three TdT^- control transfectants that have been cultured for 6 weeks following cloning. The V_H regions in the sorted subpopulations are PCR amplified and sequenced. The pie charts depict the types of mutation giving rise to V_H inactivation with the data obtained from the TdT^+ and TdT^- IgM^- subpopulations separately pooled. Abbreviations are as in Figure 4A except that "ins" indicates clones containing apparently non-templated nucleotide insertions. Clones containing deletions or duplications together with multiple nucleotide non-templated insertions are only included within the "ins" segment of the pie. Only unambiguously distinct mutational events are computed. Thus, of the 77 distinct V_H-inactivating mutations identified in the TdT^+ IgM-loss subpopulations, 30 distinct stop codon mutations are identified; if the same stop codon have been independently created within the IgM-loss population derived from a single Ramos transfectant, this would have been underscored.

The stop codons are created at variety of positions (Figure 4B) but are not randomly located. Figure 4B summarises the nature of the stop codons observed in the Rc13 and Rc14 IgM-loss populations. At least eight independent mutational events yield the nonsense mutations which account for 20 out of the 27 non-functional V_H sequences in the Rc13 database; a minimum of ten independent mutational events yield the nonsense mutations which account for 15 of the 22 non-functional V_H sequences in the Rc14 database. The numbers in parentheses after each stop codon give the number of sequences in that database that carry the relevant stop codon followed by the number of these sequences that are distinct, as discriminated on the basis of additional mutations. Analysis of stop codons in IgM-loss variants selected from four other clonal populations reveals stop codon creation at a further five locations within V_H. In data obtained in six independent experiments, stop codon creation is restricted to 16 of the 39 possible sites; the DNA sequences at these preferred sites being biased (on either coding or non-coding strand) towards the RGYW consensus.

Not surprisingly, whereas deletions and insertions account for only a small proportion of the mutations in unselected Ramos cultures (see above), they make a much greater contribution when attention is focused on V_H-inactivating mutations. It is notable that a large proportion of the IgM-loss variants can be accounted for by stop-codon/frameshift mutations in the V_H itself. This further supports the proposal that hypermutation in Ramos is preferentially targeted to the immunoglobulin V domain - certainly rather than the C domain or, indeed other genes (such as the Igα/Igβ sheath) whose mutation could lead to a surface IgM^- phenotype. It also may well be that the Ramos V_H is more frequently targeted for hypermutation than its productively rearranged V_λ, a conclusion supported by the pattern of mutations in the initial culture (Figure 1C).

Selection of cells by detection of Ig loss variants is particularly useful where those variants are capable of reverting, i.e. of reaquiring their endogenous Ig-expressing ability. The dynasty established earlier (Figure 1B) suggests not only that IgM-loss cells could arise but also that they might undergo further mutation. To confirm this, IgM-loss variants sorted from Rc13 are cloned by limiting dilution. Three weeks after cloning, the presence of IgM^+ revertants in the IgM^- subclones is screened by cytoplasmic immunofluorescence analysis of $5x10^4$ cells; their prevalence is given (Figure 4C). These IgM^+ revertants are then enriched in a single round of sorting and the V_H sequences of the clonal IgM^- variant compared to that it of its IgM^+ revertant descendants.

Cytoplasmic immunofluorescence of ten expanded clonal populations reveals the presence of IgM^+ revertants at varying prevalence (from 0.005% to 1.2%; Figure 4C) allowing a mutation rate of $1x10^{-4}$ mutations bp^{-1} $generation^{-1}$ to be calculated by fluctuation analysis. This is somewhat greater than the rate calculated by direct analysis of unselected mutations ($0.25x10^{-4}$ mutations bp^{-1} $generation^{-1}$; see above), probably in part reflecting that different IgM-loss clones revert at different rates depending upon the nature of the disrupting mutation. Indeed, the sequence surrounding the stop codons in the IgM-loss derivatives of Rc13 reveals that TAG32 conforms well to the RGYW consensus (R = purine, Y = pyrimidine and W = A or T; Rogozin and Kolchanov, 1992)

which accounts for a large proportion of intrinsic mutational hotspots (Betz et al., 1993) whereas TAA33 and TGA36 do not (Figure 4D).

Example 5: Selection of a novel Ig binding activity

In experiments designed to demonstrate development of novel binding affinities, it is noted that most members of the Ramos cell line described below express a membrane IgM molecule which binds anti-idiotype antibodies (anti-Id1 and anti-Id2), specifically raised against the Ramos surface IgM. However, a few cells retain a surface IgM, yet fail to bind the anti-idiotype antibody. This is due to an alteration in binding affinity in the surface IgM molecule, such that it no longer binds antibody. Cells which express a surface IgM yet cannot bind antibody can be selected in a single round of cell sorting according to the invention.

This is demonstrated by isolating μ positive/id-negative clones which have lost the capacity to bind to anti-Id2 despite the retention of a surface IgM, by ELISA. The clones are sequenced and in six independent clones a conserved V_H residue, K70, is found to be mutated to N, M or R as follows:

Clone	**Mutation**	
2	K70N	AAG-AAC
	S77N	AGC-AAC
4	K70M	AAG-ATG
9	S59R	AGT-AGG
	K70N	AAG-AAC
10	K70N	AAG-AAC
12	K70N	AAG-AAC
13	K70R	AAG-AGG

No mutations were observed in the light chain. Thus, it is apparent that mutants may be selected from the Ramos cell line in which the Ig molecule produced has a single base-pair variation with respect to the parent clone.

Making use of an anti-Id1, a similar population of cells is isolated which retain expression of the Igμ constant region but which have lost binding to the anti-idiotype antibody. These cells are enriched by sorting cytometry and the sequence of V_H determined (Figure 7). This reveals six mutations when compared with the consensus sequence of the starting population. Two of these mutations result in amino acid sequence changes around CDR3 (R->T at 95 and P->H at 98). Thus, selection of more subtle changes in the immunoglobulin molecule are selectable by assaying for loss of binding.

In further experiments, hypermutating cells according to the invention are washed, resuspended in PBS/BSA (10^8 cells in 0.25ml) and mixed with an equal volume of PBS/BSA containing 10% (v/v) antigen-coated magnetic beads. In the present experiment, streptavidin coated magnetic beads (Dynal) are used. After mixing at 4° C on a roller for 30 mins, the beads are washed three times with PBS/BSA, each time bringing down the beads with a magnet and removing unbound cells. remaining cells are then seeded onto 96 well plates and expanded up to 10^8 cells before undergoing a further round of selection. Multiple rounds of cell expansion (accompanied by constitutively-ongoing hypermutation) and selection are performed. After multiple rounds of selection, the proportion of cells which bind to the beads, which is initially at or close to background levels of 0.02%, begins to rise.

After 4 rounds, enrichment of streptavidin binding cells is seen. This is repeated on the fifth round (Figure 8). The low percentage recovery reflects saturation of the beads with cells since changing the cell:bead ratio from vast excess to 1:2 allows a recovery of approximately 20% from round five streptavidin binding cells (Figure 9). This demonstrates successful selection of a novel binding specificity from the hypermutating Ramos cell line, by four rounds of iterative selection.

Nucleotide sequencing of the heavy and light chains from the streptavidin binding cells predicts one amino acid change in V_H CDR3 and four changes in V_L (1 in FR1, 2 in CDR1 and 1 in CDR2) when compared with the consensus sequence of the starting population (Figure11).

To ensure that the binding of streptavidin is dependent on expression of surface immunoglobulin, immunoglobulin negative variants of the streptavidin binding cells are enriched by sorting cytometry. This markedly reduces the recovery of streptavidin binding cells with an excess of beads. The cells recovered by the Dynal-streptavidin beads from the sorted negative cells are in fact Igμ positive and most likely represent efficient recovery of Igμ streptavidin binding cells contaminating the immunoglobulin negative sorted cell population.

Preliminary data suggest that the efficiency of recovery is reduced as the concentration of streptavidin on the beads is reduced (Figure 9). This is confirmed by assaying the recovery of streptavidin binding cells with beads incubated with a range of concentrations of streptavidin (Figure 10). The percentage of cells recoverable from a binding population is dictated by the ratio of beads to cells. In this experiment the ratio is < 1: 1 beads:cells.

In a further series of experiments, a further two rounds of selection are completed, taking the total to 7. This is accomplished by reducing the concentration of streptavidin bound to the beads from 50μg/ml in round 5 to 10μg/ml in round 7. Although the secretion levels of IgM is comparable for the populations selected in rounds 4 to 7 (Figure 12), streptavidin binding as assessed by ELISA is clearly greatly increased in rounds 6 and 7, in comparison with round 4 (Figure 13).

This is confirmed by assessment of binding by Surface Plasmon Resonance on a BiaCore chip coated with streptavidin (Figure 14). The supernatant from round 7 is injected to flow across the chip at point A, and stopped at point B. At point C, anti-human IgM is injected, to demonstrate that the material bound to the streptavidin is IgM. The gradient A-B represents the association constant, and the gradient B-C to dissociation constant. From the BiaCore trace it is evident that round 6 supernatant displays superior binding characteristics to that isolated from round 4 populations or unselected Ramos cells.

Antibodies from round 6 of the selection process also show improved binding with respect to round 4. Binding of cells from round 6 selections to streptavidin-FITC aggregates, formed by preincubation of the fluorophore with a biotinylated protein, can be visualised by FACS, as shown in Figure 15. Binding to round 4 populations, unselected Ramos cells or IgM negative Ramos is not seen, indicating maturation of streptavidin binding.

Use of unaggregated streptavidin-FITC does not produce similar results, with the majority of round 6 cells not binding. This, in agreement with ELISA data, suggests that binding to streptavidin is due to avidity of the antibody binding to an array of antigen, rather than to a monovalent affinity. Higher affinity binders may be isolated by sorting for binding to non-aggregated streptavidin-FITC.

In order to determine the mutations responsible for the increased binding seen in round 6 cells over round 4 cells, the light and heavy chain antibody genes are amplified by PCR, and then sequenced. In comparison with round 4 cells, no changes in the heavy chain genes are seen, with the mutation R103S being conserved. In the light chain, mutations V23F and G24C are also conserved, but an additional mutation is present at position 46. Wild-type Ramos has an Aspartate at this position, whilst round 6 cells have an Alanine. Changes at this position are predicted to affect antigen binding, since residues in this region contribute to CDR2 of the light chain (Figure 16). It seems likely that mutation D46A is responsible for the observed increase in binding to streptavidin seen in round 6 cells.

Example 6: *In Vitro* Maturation of Ramos streptavidin binders

Ram B -> Ram C (selecting with FITC-Poly-Streptavidin)

Approximately 5 x 10^7 Ram B cells (derived from the Ramos cell line to bind Streptavidin coated microbeads) are washed with PBS and incubated on ice in 1 ml of PBS/BSA solution containing Poly-Streptavidin-FITC for 30 minutes (Poly-Streptavidin-

FITC is made by adding streptavidin FITC (20μg/ml protein content) to a biotinylated protein (10μg/ml) and incubating on ice for a few minutes prior to the addition of cells).

The cells are then washed in ice cold PBS briefly, spun down and resuspended in 500 μl PBS.

The most fluorescent 1% of cells are sorted on a MoFlo cell sorter, and this population of cells is returned to tissue culture medium, expanded to approximately $5x10^7$ cells and the procedure repeated.

After four rounds of sorting with poly-Streptavidin-FITC the cells are binding weakly to Streptavidin-FITC. Sequence of the expressed immunoglobulin V regions from this Ramos cell population reveals that amino acid number 82a in framework three of the heavy chain V region had changed from Serine to Arginine. This population of cells is called Ram C.

Ram C-> Ram D (Selecting with FITC-Streptavidin)

The next few rounds of cell sorting are done as described above but now using streptavidin-FITC (20μg/ml protein content).

After three rounds of sorting using Streptavidin-FITC the sorted cell population (called Ram D) is binding more strongly to Streptavidin FITC as assayed by FACS. Sequence of the expressed V genes reveals a further amino acid change. In framework three the amino acid at position 65, originally a Serine, has changed to Arginine

Ram D -> Ram E (Selecting with FITC-Streptavidin and unlabelled Streptavidin competition)

A subsequent sorting is done as described above using Streptavidin-FITC. However, after staining the cells on ice for 30 minutes, the cells are washed in ice cold PBS once and then resuspended in 0.5mg/ml Streptavidin and incubated on ice for 20 minutes. This is

in order to compete against the already bound Streptavidin-FITC, such that only Streptavidin-FITC that is strongly bound remains. The cells are then washed once in ice cold PBS and resuspended in 500µl PBS prior to sorting the most fluorescent 1% population as before.

After repeating this sorting protocol a further two times the Ramos cell population (Ram E) appears to bind quite strongly to Streptavidin-FITC. These cells have acquired another amino acid change in framework one of the expressed heavy chain V gene; the amino acid at position 10 had changed from Glycine to Arginine. Moreover, residue 18 has changed from Leucine to Methionine.

The results of the streptavidin maturation in Ramos cells are shown in Figure 17.

ELISA comparison

An ELISA assay performed with the supernatants of the various Ramos cell populations confirms that the IgM antibody expressed and secreted from Ramos cells has been matured *in vitro* to acquire a strong affinity for streptavidin. The results are set forth in Figure 18.

Example 7: Construction of transgene comprising hypermutation-directing sequences

It is known that certain elements of Ig gene loci are necessary for direction of hypermutation events *in vivo*. For example, the intron enhancer and matrix attachment region Ei/MAR has been demonstrated to play a critical role (Betz *et al.*, 1994). Moreover, the 3' enhancer E3' is known to be important (Goyenechea *et al.*, 1997). However, we have shown that these elements, whilst necessary, are not sufficient to direct hypermutation in a transgene.

In contrast, provision of Ei/MAR and E3' together with additional J_K-C_K intron DNA and C_K is sufficient to confer hypermutability. A βG-Cκ transgene is assembled by joining an 0.96 Kb PCR-generated KpnI-SpeI β-globin fragment (that extends from -104 with respect to the β-globin transcription start site to +863 and has artificial KpnI and SpeI restriction sites at its ends) to a subfragment of LκΔ[3'Fl] [Betz *et al.*, 1994] that extends from nucleotide 2314 in the sequence of Max *et al* [1981] through Ei/MAR, C_K and E3', and includes the 3'Fl deletion.

Hypermutation is assessed by sequencing segments of the transgene that are PCR amplified using Pfu polymerase. The amplified region extends from immediately upstream of the transcription start site to 300 nucleotides downstream of J_K5.

This chimeric transgene is well targeted for mutation with nucleotide substitutions accumulating at a frequency similar to that found in a normal Igκ transgene. This transgene is the smallest so far described that efficiently recruits hypermutation and the results indicate that multiple sequences located somewhere in the region including and flanking C_K combine to recruit hypermutation to the 5'-end of the β-globin/Igκ chimaera.

The recruitment of hypermutation can therefore be solely directed by sequences lying towards the 3'-end of the hypermutation domain. However, the 5'-border of the mutation domain in normal Ig genes in the vicinity of the promoter, some 100-200 nucleotides downstream of the transcription start site. This positioning of the 5'-border of the mutation domain with respect to the start site remains even in the βG-Cκ transgene when the β-globin gene provides both the promoter and the bulk of the mutation domain. These results are consistent with findings made with other transgenes indicating that it is the position of the promoter itself that defines the 5'-border of the mutation domain.

The simplest explanation for the way in which some if not all the κ regulatory elements contribute towards mutation recruitment is to propose that they work by bringing a hypermutation priming factor onto the transcription initiation complex. By analogy with the classic studies on enhancers as transcription regulatory elements, the Igκ enhancers

may work as regulators of hypermutation in a position and orientation-independent manner. Indeed, the data obtained with the βG-Cκ transgene together with previous results in which E3' was moved closer to C_K [Betz *et al.*, 1994] reveal that the hypermutation-enhancing activity of E3' is neither especially sensitive to its position or orientation with respect to the mutation domain.

Ei/MAR normally lies towards the 3'-end of the mutation domain. Whilst deletion of Ei/MAR drastically reduces the efficacy of mutational targeting, its restoration to a position upstream of the promoter (and therefore outside the transcribed region) gives a partial rescue of mutation but without apparently affecting the position of the 5'-border of the mutational domain. Independent confirmation of these results was obtained in transgenic mice using a second transgene, *tk-neo*::Cκ, in which a *neo* transcription unit (under control of the HSV*tk* promoter) is integrated into the C_K exon by gene targeting in embryonic stem cells [Zou, *et al.*, 1995]. In this mouse, following V_K-J_K joining, the Igκ Ei/MAR is flanked on either side by transcription domains: the V gene upstream and *tk::neo* downstream. The *tk-neo* gene is PCR amplified from sorted germinal centre B cells of mice homozygous for the *neo* insertion.

For the *tk-neo* insert in *tk-neo*::C_K mice, the amplified region extends from residues 607 to 1417 [as numbered in plasmid pMCNeo (GenBank accession U43611)], and the nucleotide sequence determined from position 629 to 1329. The mutation frequency of endogenous VJ_K rearrangements in *tk-neo* ::C_K mice is determined using a strategy similar to that described in Meyer *et al.*, 1996. Endogenous VJ_K5 rearrangements are amplified using a V_K FR3 consensus forward primer (GGACTGCAGTCAGGTTCAGTGGCAGTGGG) and an oligonucleotide LκFOR [Gonzalez-Fernandez and Milstein, (1993) PNAS (USA) 90:9862-9866] that primes back from downstream of the J_K cluster.

Although the level of mutation of the *tk-neo* is low and it is certainly less efficiently targeted for mutation than the 3'-flanking region of rearranged V_K genes in the same cell population, it appears that - as with normal V genes - the mutation domain in the *neo* gene

insert starts somewhat over 100 nucleotides downstream of the transcription start site despite the fact that Ei/MAR is upstream of the promoter.

Thus, transgenes capable of directing hypermutation in a constitutively hypermutating cell line may be constructed using Ei/MAR, E3' and regulatory elements as defined herein found downstream of J_K. Moreover, transgenes may be constructed by replacement of or insertion into endogenous V genes, as in the case of the *tk-neo* ::C_K mice, or by linkage of a desired coding sequence to the J_K intron, as in the case of the βG-Cκ transgene.

Example 8: selection of constitutively hypermutating cell line

As described above, a small proportion of V gene conversion events can lead to the generation of a non-functional Ig gene, most frequently through the introduction of frameshift mutations. Thus, the generation of sIgM loss-variants in the chicken bursal lymphoma cell line, DT40, can be used to give an initial indication of IgV gene conversion activity. Compared to the parental DT40 line, a mutant that lacks Rad54 shows a considerably diminished proportion of sIgM-loss variants (Fig. 19). A fluctuation analysis performed on multiple clones reveals that the ΔRAD54 line generates sIgM-loss variants at a frequency nearly tenfold less than that of parental DT40 whilst a ΔRAD52 line generates sIgM-loss variants at a similar frequency to wildtype cells (Fig. 19). These observations are in keeping with earlier findings concerning gene conversion in ΔRAD54 and ΔRAD52-DT40 cells (Bezzubova, *et al.*, 1997; Yamaguchi-Iwai *et al.*, 1998).

This analysis is extended to DT40 cells lacking Xrcc2 and Xrcc3. These Rad51 paralogues have been proposed to play a role in the recombination-dependent pathway of DNA damage repair (Liu *et al.*, 1998; Johnson *et al.*, 1999; Brenneman *et al.*, 2000; Takata *et al.*, 2001). Rather than giving rise to a diminished abundance of sIgM-loss variants, the ΔXRCC2 and ΔXRCC3 lines show a much greater accumulation of loss variants than the parental line (Fig. 19). In the case of ΔXRCC2-DT40, transfection of the human Xrcc2 cDNA under control of the human β-globin promoter causes the frequency of generation of sIgM-loss variants to revert to close to wildtype values. Figure 19 shows the generation of sIgM-loss variants by wildtype and repair-deficient DT40 cells. Flow

cytometric analyses of the heterogeneity of sIgM expression in cultures derived by 1 month of clonal expansion of single $sIgM^+$ normal (WT) or repair-deficient (ΔRAD54, ΔRAD52, ΔXRCC2, ΔXRCC3) DT40 cells are shown in panel (a). An analysis of cultures derived from three representative $sIgM^+$ precursor clones is shown for each type of repair-deficient DT40. The percentage of $sIgM^-$ cells in each analysis is indicated with the fluorescence gate set as eightfold below the centre of the $sIgM^+$ peak. Panel (b) shows fluctuation analysis of the frequency of generation of sIgM-loss variants. The abundance of sIgM-loss variants is determined in multiple parallel cultures derived from $sIgM^+$ single cells after 1 month of clonal expansion; median percentages are noted above each data set and indicated by the dashed bar. The [pβG-hXRCC2]ΔXRCC2 transfectants analysed are generated by transfection of pβG-hXRCC2 into $sIgM^+$ DT40-ΔXRCC2 subclones that have 6.4% and 10.2% $sIgM^-$ cells in the fluctuation analysis. The whole analysis is performed on multiple, independent $sIgM^+$ clones (with distinct, though similar ancestral Vλ sequences) giving, for each repair-deficient line, average median frequencies at which sIgM-loss variants are generated after 1 month of WT (0.4%), ΔRAD54 (0.07%), ΔRAD52 (0.4%), ΔXRCC2 (6%) and ΔXCRCC3 (2%).

Since deficiency in both Xrcc2 and Xrcc3 is associated with chromosomal instability (Liu *et al.*, 1998); Cui *et al.*, 1999; Deans *et al.*, 2000; Griffin *et al.*, 2000), it is possible that the increased frequency of sIgM-loss variants could reflect gross rearrangements or deletions within Ig loci. However, Southern blot analysis of 24 $sIgM^-$ subclones of ΔXRCC3-DT40 does not reveal any loss or alteration of the 6kb SalI-BamHI fragment containing the rearranged Vλ.

Therefore, to ascertain whether more localised mutations in the V gene could account for the loss of sIgM expression, the rearranged Vλ segments in populations of $sIgM^-$ cells that are sorted from wildtype, ΔXRCC2- and ΔXRCC3-DT40 subclones after one month of expansion are cloned and sequenced.

Cell culture, transfection and analysis

DT40 subclone CL18 and mutants thereof are propagated in RPMI 1640 supplemented with 7% foetal calf serum, 3% chicken serum (Life Technologies), 50μM 2-

mercaptoethanol, penicillin and streptomycin at 37°C in 10% CO_2. Cell density was maintained at between $0.2 - 1.0 \times 10^6$ ml^{-1} by splitting the cultures daily. The generation of the DT40 derivatives carrying targeted gene disruptions has been described elsewhere (Bezzubova *et al.*, 1997; Yamaguchi-Iwai *et al.*, 1998; Takata *et al.*, 1998, 2000, 2001). Transfectants of ΔXRCC2-DT40 harbouring a pSV2-neo based plasmid that contains the XRCC2 open reading frame (cloned from HeLa cDNA) under control of the β-globin promoter are generated by electroporation.

CL18 is an $sIgM^-$ subclone of DT40 and is the parental clone for the DNA repair-mutants described here. Multiple $sIgM^+$ subclones are obtained from both wild type and repair-deficient mutants using a Mo-Flo (Cytomation) sorter after staining with FITC-conjugated goat anti-chicken IgM (Bethyl Laboratories). There is little variation in the initial Vλ sequence expressed by all the $sIgM^+$ DT40-CL18 derived repair-deficient cells used in this work since nearly all the $sIgM^+$ derivatives have reverted the original CL18 Vλ frameshift by gene conversion using the ψV8 donor (which is most closely related to the frameshifted CL18 CDR1).

Mutation analysis

Genomic DNA is PCR amplified from 5000 cell equivalents using Pfu Turbo (Stratagene) polymerase and a hotstart touchdown PCR [8 cycles @ 95°C 1', 68-60°C (at 1°C per cycle) 1', 72 °C 1'30"; 22 cycles @ 94°C 30", 60 °C 1', 72°C 1'30"]. The rearranged Vλ is amplified using CVLF6 (5'-CAGGAGCTCGCGGGGCCGTCACTGATTGCCG; priming in the leader-Vλ intron) and CVLR3 (5'-GCGCAAGCTTCCCCAGCCTGCCGCCAAGTCCAAG; priming back from 3' of Jλ); the unrearranged Vλ1 using CVLF6 with CVLURR1 (5'-GGAATTCTCAGTGGGAGCAGGAGCAG); the rearranged VH gene using CVH1F1 (5'-CGGGAGCTCCGTCAGCGCTCTCTGTCC) with CJH1R1 (5'-GGGGTACCCGGAGGAGACGATGACTTCGG) and the Cλ region using CJCIR1F (5'-GCAGTTCAAGAATTCCTCGCTGG; priming from within the Jλ-Cλ intron) with CCMUCLAR (5'-GGAGCCATCGATCACCCAATCCAC; priming back from within Cλ). After purification on QIAquick spin columns (Qiagen), PCR products are cut with the appropriate restriction enzymes, cloned into pBluescriptSK and sequenced using the

T3 or T7 primers and an ABI377 sequencer (Applied Biosystems). Sequence alignment (Bonfield *et al.*, 1995) with GAP4 allowed identification of changes from the consensus sequence of each clone.

All sequence changes are assigned to one of three categories: gene conversion, point mutation or an ambiguous category. This discrimination rests on the published sequences of the Vλ pseudogenes that could act as donors for gene conversion. The database of such donor sequences is taken from Reynaud et al. (1987) but implementing the modifications (McCormack *et al.*, 1993) pertaining to the Igλ G4 allele appropriate (Kim *et al.* 1990) to the expressed Igλ in DT40. (The sequences/gene conversions identified in this work supported the validity of this ψVλ sequence database). For each mutation the database of Vλ pseudogenes is searched for potential donors. If no pseudogene donor containing a string ≥9bp could be found then it is categorised as an untemplated point mutation. If a such a string is identified and there are further mutations which could be explained by the same donor, then all these mutations are assigned to a single gene conversion event. If there are no further mutations then the isolated mutation could have arisen through a conversion mechanism or could have been untemplated and is therefore categorised as ambiguous.

With regard to the Vλ sequences cloned from the sIgM$^-$ subpopulations sorted from multiple wildtype DT40 clones, 67% carry mutations: in the majority (73%) of cases, these mutations render the Vλ obviously non-functional, as shown in Figure 20. Presumably, most of the remaining sIgM$^-$ cells carry inactivating mutations either in V_H or outside the sequenced region of Vλ. Figure 20 shows analyses of Vλ sequences cloned from sIgM-loss variants. In panel (a), comparison of Vλ sequences obtained from sIgM-loss cells that have been sorted from parental sIgM$^+$ clones of normal or Xrcc2-deficient DT40 cells after 1 month of clonal expansion. Each horizontal line represents the rearranged Vλ1/Jλ (427 bp) with mutations classified as described above as point mutations (lollipop), gene conversion tracts (horizontal bar above line) or single nucleotide substitutions which could be a result of point mutation or gene conversion (ambiguous, vertical bar). Hollow boxes straddling the line depict deletions, triangles indicate a duplications. Pie charts are shown in panel (b), depicting the proportion of Vλ

sequences that carry different numbers of point mutations (PM), gene conversions (GC) or mutations of ambiguous origin (Amb) amongst sorted sIgM-loss populations derived from wildtype, ΔXRCC2 or ΔXRCC3 DT40 $sIgM^+$ clones after 1 month of clonal expansion. The sizes of the segments are proportional to the number of sequences carrying the number of mutations indicated around the periphery of the pie. The total number of Vλ sequences analysed is indicated in the centre of each pie with the data compiled from analysis of four subclones of wildtype DT40, two of ΔXRCC2-DT40 and three of ΔXRCC3-DT40. Deletions, duplications and insertions are excluded from this analysis; in wildtype cells, there are additionally 6 deletions, 1 duplication and 1 insertion. There are no other events in ΔXRCC2-DT40 and a single example each of a 1bp deletion and a 1bp insertion in the ΔXRCC3-DT40 database.

Causes of Vλ gene inactivation in wildtype, ΔXRCC2 (ΔX2) and ΔXRCC3 (ΔX3) DT40 cells expressed as a percentage of the total sequences that contained an identified inactivating mutation are set forth in panel (c): Missense mutation (black); Gene conversion-associated frameshift (white); Deletions, insertions or duplication-associated frameshift (grey). Additional mutational events associated with each inactivating mutation are then shown in (d). The data are expressed as the mean number of additional mutations associated with each inactivating mutation with the type of additional mutation indicated as in panel (c). Thus, ΔXRCC2-DT40 has a mean of 1.2 additional point mutations in addition to the index inactivating mutation whereas wildtype DT40 has only 0.07.

As detailed above, the mutations may be classified as being attributable to gene conversion templated by an upstream Vλ pseudogene, to non-templated point mutations or as falling into an ambiguous category. Most (67%) of the inactivating mutations are due to gene conversion although some (15%) are stop codons generated by non-templated point mutations demonstrating that the low frequency of point mutations seen here and elsewhere (Buerstedde *et al.*, (1985); Kim *et al.*, 1990) in DT40 cells is not a PCR artefact but rather reveals that a low frequency of point mutation does indeed accompany gene conversion in wildtype DT40.

A strikingly different pattern of mutation is seen in the Vλ sequences of the sIgM-loss variants from ΔXRCC2-DT40. Nearly all the sequences carry point mutations, typically with multiple point mutations per sequence. A substantial shift towards point mutations is also seen in the sequences from the sIgM⁻ ΔXRCC3-DT40 cells. Thus, whereas a Vλ-inactivating mutation in wild type DT40 is most likely to reflect an out of frame gene conversion tract, in ΔXRCC2/3 it is likely to be a missense mutation (Fig. 20c). Furthermore, whereas most of the nonfunctional Vλ sequences obtained from sorted sIgM-loss variants of ΔXRCC2-DT40 (53%) or ΔXRCC3-DT40 (64%) carry additional point mutations in addition to the Vλ-inactivating mutation, such hitchhiking is only rarely observed in the nonfunctional Vλ sequences from the parental DT40 line (7%; Fig. 20d).

All these observations suggest that the high prevalence of sIgM-loss variants in ΔXRCC2/3-DT40 cells simply reflects a very high frequency of spontaneous IgV gene hypermutation in these cells. Figure 21 represents analyses of Ig sequences cloned from unsorted DT40 populations after one month of clonal expansion. The Vλ sequences obtained from representative, wildtype and ΔXRCC2 DT40 clones are presented in panel (a) with symbols as in Fig. 20. In panel (b), pie charts are shown depicting the proportion of the Vλ sequences carrying different numbers of the various types of mutation as indicated. The data are pooled from analysis of independent clones: wildtype (two clones), ΔXRCC2 (four clones) and ΔXRCC3 (two clones). In addition to the mutations shown, one ΔXRCC2-DT40 sequence contained a 2bp insertion in the leader intron which was not obviously templated from a donor pseudogene and one ΔXRCC3-DT40 sequence carried a single base pair deletion also in the leader intron.

Mutation at other loci of ΔXRCC2-DT40 is shown in panel (c). Pie charts depict the proportion of sequences derived from 1 month-expanded ΔXRCC2-DT40 cells that carry mutations in the rearranged V_H (272bp extending from CDR1 to the end of J_H) of the rearranged heavy chain of, in the unrearranged Vλ1 on the excluded allele (458 bp) and in the vicinity of Cλ (425bp extending from the Jλ-Cλ intron into the first 132bp of Cλ). Analysis of known V_H pseudogene sequences (Reynaud *et al.*, 1989) does not indicate that any of the mutations observed in the rearranged V_H are due to gene conversion,

strongly suggesting that they are due to point mutation although this assignment cannot be regarded as wholly definitive. The mutation prevalences in these data sets are: 1.6×10^{-3} mutations.bp^{-1} for V_H, 0.03×10^{-3} for the unrearranged Vλ1 and 0.13×10^{-3} for Cλ as compared to 2.0×10^{-3} for point mutations in the rearranged Vλ1 in ΔXRCC2-DT40, 0.13×10^{-3} for point mutations in rearranged Vλ1 in wildtype DT40 and 0.04×10^{-3} for background PCR error.

The distribution of point mutations across Vλ1 is shown in panel (d). The ΔXRCC2-DT40 consensus is indicated in upper case with the first base corresponding to the 76th base pair of the leader intron. Variations found in the ΔXRCC3-DT40 consensus are indicated in italic capitals below. The mutations are shown in lower case letters above the consensus with those from ΔXRCC2-DT40 in black and those from ΔXRCC3-DT40 in mid-grey. All mutations falling into the point mutation and ambiguous categories are included. Correction has been made for clonal expansion as described previously (Takata *et al.*, 1998) so each lower case letter represents an independent mutational event. The majority of the 27 mutations thereby removed from the original database of 158 are at one of the seven major hotspots; the correction for clonality will, if it gives rise to any distortion, lead to a underestimate of hotspot dominance. Of the seven major hotspots (identified by an accumulation of ≥ 5 mutations), five conform to the AGY consensus sequence on one of the two strands as indicated with black boxes. Nucleotide substitution preferences (given as a percentage of the database of 131 independent events) as shown in panel (e) are deduced from the point mutations in sequences from unselected ΔXRCC2- and ΔXRCC3-DT40. A similar pattern of preferences is evident if the ΔXRCC2/ΔXRCC3 databases are analysed individually.

The spontaneous Vλ mutation frequency in wildtype and ΔXRCC2/3-DT40 cells is analysed by PCR amplifying the rearranged Vλ segments from total (unsorted) DT40 populations that have been expanded for 1 month following subcloning. The result reveals that there is indeed a much higher spontaneous accumulation of mutations in the ΔXRCC2 and ΔXRCC3 cells than in the parental DT40 (Fig. 21a, b). In ΔXRCC2-DT40 cells, mutations accumulate in Vλ at a rate of about 0.4×10^{-4} bp^{-1}.$generation^{-1}$ (given an

approximately 12 hour division time), a value similar to that seen in the constitutively mutating human Burkitt lymphoma line Ramos.

Somatic hypermutation in germinal centre B cells in man and mouse is preferentially targeted to the rearranged immunoglobulin V_H and V_L segments. A similar situation applies to the point mutations in ΔXRCC2-DT40 cells. Thus, a significant level of apparent point mutation is also seen in the productively rearranged V_H1 gene (Fig. 3c). However, this does not reflect a general mutator phenotype since mutation accumulation is much lower in Cλ than in the rearranged Vλ and is also low in the unrearranged Vλ on the excluded allele where the apparent mutation rate does not rise above the background level ascribable to the PCR amplification itself (Fig. 21c).

The distribution of the mutations over the Vλ domain in ΔXRCC2-DT40 cells is strikingly non-random. The mutations, which are predominantly single nucleotide substitutions, show preferential accumulation at hotspots that conform to an AGY (Y=pyrimidine) consensus on one of the two DNA strands (Fig. 21d). They also occur overwhelmingly (96%) at G/C. This G/C-biased, hotspot-focused hypermutation in ΔXRCC2-DT40 cells, although exhibiting somewhat less of a bias in favour of nucleotide transitions, is strikingly similar to the pattern of V gene hypermutation described in cultured human Burkitt lymphoma cells as well as that occurring in vivo in frog, shark and Msh2-deficient mice (Rada *et al.*, 1998; Diaz *et al.*, 2001). The IgV gene hypermutation that occurs in vivo in man and normal mice appears, as previously discussed, to be achieved by this hotspot-focused G/C biased component acting in concert with a mechanism that targets A/T (Fig. 21e).

Thus, whereas the DT40 chicken bursal lymphoma line normally exhibits a low frequency of IgV diversification by gene conversion, a high frequency of constitutive IgV gene somatic mutation (similar in nature to that occurring in human B cell lymphoma models) can be elicited by ablating Xrcc2 or Xrcc3. This provides strong support to the earlier proposal that IgV gene conversion and hypermutation might constitute different ways of resolving a common DNA lesion (Maizels *et al.*, 1995; Weill *et al.*, 1996). Recent data suggest that the initiating lesion could well be a double strand break (Sale & Neuberger,

1998; Papavasilou *et al.*, 2000; Bross *et al.* 2000) and it would therefore appear significant that both Xrcc2 and Xrcc3 have been implicated in a recombination-dependent pathway of DNA break repair (Liu *et al.*, 1998; Johnson *et al.*, 1999; Pierce *et al.*, 1999; Brennerman *et al.*, 2000; Takata *et al.*, 2001). Indeed, a similar induction of IgV gene hypermutation in DT40 cells is achieved by ablating another gene (RAD51B) whose product is implicated in recombination-dependent repair of breaks (Takata *et al.*, 2000) but not by ablating genes for Ku70 and DNA-PK_{cs} which are involved in non-homologous end-joining. Figure 22 shows the analysis of sIgM-loss variants in DT40 cells deficient in DNA-PK, Ku70 and Rad51B. Fluctuation analysis of the frequency of generation of sIgM-loss variants after 1 month of clonal expansion is shown in panel (a). The median values obtained with wildtype and ΔXRCC2 DT40 are included for comparison. Pie charts depicting the proportion of Vλ sequences amplified from the sIgM-loss variants derived from two $sIgM^+$ Rad51B-deficient DT40 clones that carry various types of mutation as indicated are shown in panel (b). In addition, one sequence carried a 9 bp deletion, one carried a 4 bp duplication and one carried a single base pair insertion.

The results, however, do not simply suggest that, in the absence of Xrcc2, a lesion which would normally be resolved by gene conversion is instead resolved by a process leading to somatic hypermutation. First, ΔXRCC2-DT40 cells retain the ability to perform IgV gene conversion, albeit at a somewhat reduced level (Fig. 21b). Second, the frequency of hypermutation in ΔXRCC2-DT40 cells is about an order of magnitude greater than the frequency of gene conversion in the parental DT40 line. It is therefore likely that, in normal DT40 cells, only a minor proportion of the lesions in the IgV gene are subjected to templated repair from an upstream pseudogene thereby leading to the gene conversion events observed. We believe that the major proportion of the lesions are subjected to a recombinational repair using the identical V gene located on the sister chromatid as template and which is therefore 'invisible'. This would be consistent with the observations of Papavasiliou and Schatz (2000) who found that detectable IgV gene breaks in hypermutating mammalian B cells are restricted to the G2/S phase. In the absence of Xrcc2, Xrcc3 or Rad51B, we propose that the 'invisible' sister chromatid-dependent recombinational repair is perverted, resulting in hypermutation. Whether this hypermutation reflects that the sister chromatid-dependent recombinational repair

becomes error-prone in the absence of Xrcc2/3 or whether it reflects an inhibition of such repair thereby revealing an alternate, non-templated mechanism of break resolution is an issue that needs to be addressed. This question is not only important for an understanding of the mechanism of hypermutation but may also provide insight into the physiological function of the Rad51 paralogues.

Example 9: Affinity maturation in Δxrcc2 DT40 IgM

A population of Xrcc2-deficient DT40 cells which had been expanded for several months was used to determine whether the action of hypermutation on the unique V_HDJ_H/V_LJ_L rearrangement in these cells could generate sufficient functional diversity to allow the evolution of maturing lineages of antibodies to a significant proportion of antigens tested. We used two methods for selection. In a first approach, cells were incubated with soluble aggregates formed by mixing FITC-streptavidin with different biotinylated antigens (casein, insulin, ovalbumin (Ova), thyroglobulin (Tg) and a rat monoclonal antibody (Ab)); binding variants were then enriched by flow cytometry. Alternatively, variants were selected using magnetic beads coated with *S. aureus* Protein A or human serum albumin (HSA). None of the antigens tested showed detectable binding to the parental DT40 or its secreted IgM. After six sequential rounds of selection, there were essentially three types of outcome. In the case of insulin and casein, there was no evidence for enrichment of binding variants. In the cases of the rat monoclonal antibody and Protein A, specific binders were obtained. And in the cases of HSA, Ova and Tg, binders were obtained but these exhibited varying degrees of polyreactivity. Thus, with the selection performed using an aggregate of biotinylated rat IgG mAb S7 with FITC-streptavidin (Fig. 23), specific binding was already evident by round three. Subsequent enrichments were performed using the rat IgG S7 mAb directly conjugated to phycoerythrin (PE). The cells obtained in round six (DT-Ab6) were specific for the rat IgG S7 mAb as judged by the lack of staining by a variety of other reagents. ELISA of the culture supernatant demonstrated that the binding of S7 mAb by DT-Ab6 cells was conferred by the DT-Ab6 IgM itself (Fig. 23b). This DT-Ab6 IgM is in fact an anti-S7 idiotype. It recognises purified S7 mAb (and can be used for staining permeabilised S7 hybridoma cells) but this interaction is not competable by other rat immunoglobulins (Fig.

23c, d). DT-Ab6 cells could be stained using high dilutions of PES7 mAb conjugate - suggesting a high affinity of interaction. By performing the staining at a fixed ratio of molecules PE-S7 mAb: DT-Ab6 cells but varying the volume, an affinity in the range of 5.8 nM was deduced (Fig. 23f).

Sequence analysis revealed that DT-Ab6 V_H/V_L carry a total of 19 amino acid substitutions compared to the parental DT40 sequence (Fig. 23g).

In the case of the bead selections performed using streptavidin beads coated with biotinylated Protein A, binding variants were evident by round two (DT-P2; Fig. 24a). However, many of the cells in the DT-P2 population were sticky in that they also bound to various other types of bead that displayed neither streptavidin nor Protein A. Further enrichments using tosylated beads to which Protein A had been directly conjugated yielded a population of cells (DT-P4) that could be stained with a FITC conjugated Protein A aggregate (Fig. 24b). Serial selection for binding to this aggregate by use of flow cytometry gave rise to a population (DT-P9) which could be stained weakly with unaggregated FITC-Protein A; further sorting with FITC-Protein A then gave rise to more brightly staining descendants (Fig. 24b). The IgM secreted by DT-P14 cells (as well as by several of its precursors) bound well to Protein A as judged by both ELISA and immunoprecipitation (Fig. 24c, d). Direct binding assays using radiolabelled Protein A indicated a substantial increase in affinity had occurred between DT-P9 and DT-P14 (Fig. 24e), consistent with the flow cytometric analysis (Fig.24b). This increase, which is entirely due to a Ala->Val substitution adjacent to V_H CDR2 (Fig. 24h), yields an IgM with an apparent affinity for Protein A of about 0.32 nM (Fig. 24f). Interestingly, the interaction between DT-P IgM and Protein A differs from the well characterised interaction between Protein A and the Fc portion of rabbit IgG not just in the fact that the DT-P/Protein A interaction is with the V rather than C portion of the IgM molecule (mutations are found in V_H not in C_μ) but also in the fact that it is likely that different sites on Protein A are used to interact with DT-P IgM and rabbit IgG since high concentrations of rabbit IgG do not inhibit the staining of DT-P4/P9 cells by FITC-Protein A (Fig. 24g). Indeed, an enhancement of staining is seen, presumably due to aggregation of Protein A by the rabbit IgG.

The sequence of the DT-P14 V_H/V_L reveals 17 amino acid substitutions compared to the parental DT40 sequence. Five of these mutations are shared with the rat idiotype-specific variant DT-Ab6. However, at least four of the five mutations must have occurred independently in the two dynasties (rather than reflecting descent from a common mutated precursor in the pool of Xrcc2-deficient DT40 cells) since these four mutations common to DT-P14 and DT-Ab6 are not found in DT-P4. This tendency to repeat substitutions might in part reflect that mutation in DT40 is largely restricted to the hotspot-focussed GC-biased first phase of mutations (lacking some of the breadth of mutation that we have ascribed to the A/T-biased second phase; Sale and Neuberger, 1998; Rada *et al.*, 1998) although it is also possible that they confer an advantage by predisposing the antibody structure to maturability. It is notable that whilst Xrcc2-deficient DT40 cells retain the ability to perform IgV gene conversion (albeit at much lower frequency than somatic hypermutation)5, comparison of the V gene sequences in DT-P14 and DT-Ab6 cells (Fig. 24h) with those of the germline V segments (Reynaud *et al.*, 1987; 1989) reveals that the changes are largely (if not exclusively) due to point mutations rather than gene conversions. The cells obtained after six rounds of selection using HSA-derivatised carboxylated magnetic beads (DT-H6) stained not only with FITC-HSA, but also with FITC-Tg, FITC-Ova and Cychrome-conjugated streptavidin despite the fact that the DT-H6 cell population had never been exposed to these other antigens; further selections with HSA simply increased the brightness of polyspecific staining (Fig. 25a). A similar, though distinct, pattern of polyspecificity was evident in the cell population that had been subjected to flow cytometric enrichment using complexes of FITC-streptavidin with either biotinylated Tg or biotinylated Ova (Fig 25a, d). Analysis of subpopulations revealed that the apparent polyspecificity is not a reflection of cellular heterogeneity. The polyspecific staining was mediated by the surface IgM itself since sIgM-loss variants lose antigen-binding activity. Furthermore, antigen binding is readily detected by ELISA of the culture supernatants (Fig. 25b, c).

Polyreactive antibodies are well described in man and mouse (both in serum and amongst hybridomas), where the issue has been raised as to whether they constitute a good starting point for the evolution of monospecificity (Casali and Schettino, 1996; Bouvet and

Dighiero, 1998). The same issue arises in the *in vitro* selection system (though without the constraint of avoiding autoimmunity). We therefore tested whether it is possible to evolve the DT-O6 population towards increased specificity for Ova by flow cytometric enrichment for FITC-Ova$_{bright}$ /Cychrome-streptavidin$_{dull}$ cells. After multiple rounds of sorting, cells displaying greater specificity for FITC-Ova were obtained (Fig. 25d). It will be interesting to ascertain whether a polyspecific binding population provides a better starting point for the evolution of specificity than the parental DT40 population.

The results presented here clearly demonstrate that hypermutating cell lines can be used for both the derivation and iterative maturation of antibodies *in vitro*. Given the genetic tractability of DT40, it is possible to extend the application to transfected IgV genes and thereby mature the affinity of existing antibodies. Furthermore, since the hypermutation mechanism can target heterologous genes put in place of the rearranged IgV segment (Yelamos *et al.*, 1995), it may well prove possible to extend the strategy to the maturation of other ligand/receptor pairs. With regard to the *de novo* selection of antibodies, it is striking that the action of somatic hypermutation on a single V_H/V_L rearrangement has generated a repertoire from which it has been possible to select and mature high affinity binders to two of the seven antigens tested as well as obtain signs of initial low affinity (but maturable) binding to three of the others. This reflects the fact that very low antigen affinities suffice to initiate the selection: these binding sites are then maturable. Clearly, if wishing to extrapolate the approach to allow the *in vitro* production of high-affinity human monoclonal antibodies, it would be advantageous to exploit the genetic tractability of DT40 so as to generate a primary repertoire that includes more than a single V_H/V_L rearrangement. The results obtained here indicate that this primary repertoire of rearrangements would be orders of magnitude smaller than the primary repertoire used *in vivo*.

Example 10: Isolation of naturally-occuring constitutively hypermutating EBV positive BL cell lines

A survey of naturally occurring EBV$^+$ BL cell lines revealed an absence of a clearly identifiable population of sIgM-loss variants amongst many of them (e.g. Akata, BL74,

Chep, Daudi, Raji, and Wan). However, a clear sIgM$^{-/low}$ population was noted in two of these EBV$^+$ cell lines, ELI-BL and BL16, suggesting an intrinsic hypermutation capacity. sIgM expression profiles of Ramos, EHRB, ELI-BL, and BL16 are shown in Fig.26a. The sIgM$^{-/low}$ cell population is boxed and the percentage of cells therein indicated. Each dot represents one cell. Note that the sizable sIgM$^{-/low}$ population in BL16 is in part due to less intensely staining positive cells, which also occluded fluctuation analyses. ELI-BL harbors a type 2 EBV, resembles germinal center B cells, and expresses a latency gene repertoire consisting only of EBNA1 and the non-coding EBER and Bam A RNAs (Rowe, *et al.*,1987) BL16 also contains a type 2 virus but, in contrast to ELI-BL, it appears more LCL-like and expresses a full latency gene repertoire (Rooney *et al.*, 1984; Rowe *et al.*, 1987).

Although a clear sIgM$^{-/low}$ population was visible in ELI-BL and BL16 cultures, it was important to address whether these variants could be attributed to bonafide hypermutation. This was assessed by fluctiation analysis. In brief, (sub)clones were transferred to 24 or 48 well plates, maintained with fresh medium for 3 to 8 weeks, and analyzed by washing cells (1-2.5 x 10^5) twice in PBS/3% FBS, staining (30 min on ice) with the relevant antibody or antibody combination (below) and again washing prior to analysis of at least 10^4 cells by flow cytometry (FACSCalibur, Becton Dickinson). Antibodies used were R-phycoerythrin-conjugated, goat anti-human IgM (μ-chain specific; Sigma), fluorescein isothiocyanate (FITC)-conjugated, mouse monoclonal anti-Ramos idiotype [ZL16/1 (Zhang *et al.*,1995); provided generously by M. Cragg and M. J. Glennie, Tenovus Research Laboratory, Southampton], and FITC-conjugated, goat anti-mouse IgM (Southern Biotechnology Associates, Inc.). Data were acquired and analyzed using CellQuest software (Becton Dickinson).

Unless noted otherwise, cells compared in fluctuation analyses were derived, cultured, and analyzed in parallel. The median (as opposed to the mean) percentage of sIgM-loss variants amongst a number of identically-derived (sub)clones is used as an indicator of a cells somatic hypermutation capacity to minimize the effects of early mutational events Fluctuation analysis of ELI-BL subclones revealed that the sIgM$^{-/low}$ variants were indeed being generated at high frequency during in vitro culture (Fig. 26b; each cross represents the percentage of cells falling within the sIgM$^{-/low}$ window following a 1 month outgrowth of a single subclone; the median percentages are indicated), and V_H

sequence analysis, in the case of BL16 subclones, confirmed that this instability reflected somatic hypermutation (Fig. 26c). Base substitution mutations are indicated in lower case letters above the 338 bp consensus DNA sequence in triplets of capital letters. Complementarity-determining regions and partial PCR primer sequences are underlined and emboldened, respectively. The corresponding amino acid sequence is indicated by single capital letters. This consensus sequence differs at two positions from GenBank entry gi.2253343 [TCA (Ser20) - TCT and AGC (Ser55) - ACC (Thr)].

Considerable V_H sequence diversity, including several sequences with multiple base substitution mutations, and an overall high V_H mutation frequency indicated that hypermutation is ongoing in BL16. Moreover, despite the relatively small number of V_H sequences sampled, one dynastic relationship could be inferred [1st mutation at Gly54 (GGT - GAT); 2nd mutation at Val92 (GTG - ATG)]. Finally, like Ramos, most of the BL16 V_H base substitution mutations occurred at G or C nucleotides (24/33 or 73%) and clustered within the complementarity determining regions (underlined in Fig. 26c). Thus, several hallmarks of ongoing hypermutation were also distinguishable in two natural EBV^+ BL cell lines, one expressing a limited latency gene repertoire and the other expressing a full combination. It was therefore clear that somatic hypermutation can proceed unabated even in the presence of EBV.

References

Adetugbo, K., Milstein, C. and Secher, D.S. (1977). Molecular analysis of spontaneous somatic mutants. Nature 265, 299-304.

Bahler, D.W. and Levy, R. (1992). Clonal evolution of a follicular lymphoma: Evidence for antigen selection. Proc. Natl. Acad. Sci. USA 89, 6770-6774.

Bass, S., R. Greene, and J.A. Wells. (1990). Hormone Phage: An Enrichment Method for Variant Proteins With Altered Binding Properties. *Proteins*. 8, 309-314.

Berek, C. and Milstein, C. (1988). The dynamic nature of the antibody repertoire. Immunol. Rev. 105, 5-26.

Betz, A. G., Milstein, C., Gonzalez-Fernandez, A., Pannell, R., Larson, T. and Neuberger, M. S., *Cell* 1994. *77*: 239-248.

Betz, A.G., Neuberger, M.S. and Milstein, C. (1993). Discriminating intrinsic and antigen-selected mutational hotspots in immunoglobulin V genes. Immunol. Today 14, 405-411.

Bezzubova, O., Silbergleit, A., Yamaguchi-Iwai, Y., Takeda, S. & Buerstedde, J. M. Reduced X-ray resistance and homologous recombination frequencies in a RAD54-/- mutant of the chicken DT40 cell line. *Cell* **89**, 185-193. (1997).

Bonfield, J.K., Smith, K.F. and Staden, R. (1995). A new DNA-sequence assembly program. Nucleic Acids Res. 23, 4992-99.

Bouvet J.P. & Dighiero, G. From natural polyreactive autoantibodies to a la carte monoreactive antibodies to infectious agents: is it a small world after all? *Infect. Immun.* **66**, 1-4 (1998).

Braeuninger, A., Küppers, R., Strickler, J.G., Wacker, H.-H., Rajewsky, K. and Hansmann, M.-L. (1997). Hodgkin and Reed-Sternberg cells in lymphocyte predominant disease represent clonal populations of germinal center-derived tumor B cells. Proc. Natl. Acad. Sci. USA 94, 9337-9342.

Brenneman, M. A., Weiss, A. E., Nickoloff, J. A. & Chen, D. J. XRCC3 is required for efficient repair of chromosome breaks by homologous recombination. *Mutat Res* 459, 89-97. (2000).

Bross, L., Fukita, Y., McBlane, F., Demolliere, C., Rajewsky, K. & Jacobs, H. DNA double-strand breaks in immunoglobulin genes undergoing somatic hypermutation. *Immunity* **13**, 589-597. (2000)

Brüggemann, M., Radbruch, A. and Rajewsky, K. (1982). Immunoglobulin V region variants in hybridoma cells. I. Isolation of a variant with altered idiotypic and antigen binding specificity. EMBO J. 1, 629-634.

Buerstedde, J. M., Reynaud, C.A., Humphries, E.H., Olson, W., Ewert, D.L. & Weill, J.C. Light chain gene conversion continues at high rate in an ALV-induced cell line. *Embo J* **9**, 921-927. (1990)

Capizzi and Jameson, (1973) Mutat. Res. 17:147-8

Casali P. & Schettino, E.W. Structure and function of natural antibodies. *Curr. Top. Microbiol. Immunol.* **10**, 167-179 (1996).

Chapman, C.J., Mockridge, C.I., Rowe, M., Rickinson, A.B. and Stevenson, F.K. (1995). Analysis of VH genes used by neoplastic B cells in endemic Burkitt's lymphoma shows somatic hypermutation and intraclonal heterogeneity. Blood 85, 2176-2181.

Chapman, K.B. and Szostak, J.W. (1994) Curr. op. Struct. Biol., 4, 618-622.

Chui, Y.-L., Lozano, F., Jarvis, J.M., Pannell, R. and Milstein, C. (1995). A reporter gene to analyse the hypermutation of immunoglobulin genes. J. Mol. Biol. 249, 555-563.

Clackson, T. and Wells, J.A. (1994) Trends Biotechnol, 12, 173-84.

Coffino, P. and Scharff, M.D. (1971). Rate of somatic mutation in immunoglobulin production by mouse myeloma cells. Proc. Natl. Acad. Sci. USA 68, 219-223.

Cui, X., Brenneman, M., Meyne, J., Oshimura, M., Goodwin, E.H. & Chen D.J. The XRCC2 and XRCC3 repair genes are required for chromosome stability in mammalian cells. *Mutat Res* 434, 75-88. (1999).

Cull, M.G., Miller, J.F. and Schatz, P.J. (1992) Proc Natl Acad Sci U S A, 89, 1865-9.

Deans, B., Griffin, C. S., Maconochie, M. & Thacker, J. Xrcc2 is required for genetic stability, embryonic neurogenesis and viability in mice. *Embo J* 19, 6675-6685. (2000).

Denépoux, S., Razanajaona, D., Blanchard, D., Meffre, G., Capra, J.D., Banchereau, J and Lebecque, S. (1997). Induction of somatic mutation in a human B cell line in vitro. Immunity 6, 35-46.

Diaz, M., Flajnik, M. F. & Klinman, N. Evolution and the molecular basis of somatic hypermutation of antigen receptor genes. *Philos Trans R Soc Lond B Biol Sci* 356, 67-72. (2001)

Drake, J.W. (1998) Genetics 148:1667-1686.

Ellington, A.D. and Szostak, J.W. (1990) Nature, 346, 81822.

Gold, L., Polisky, B., Uhlenbeck, 0. and Yarus, M. (1995) Annu Rev Biochem, 64, 763-97.

Goossens, T., Klein, U. and Küppers, R. (1998). Frequent occurrence of deletions and duplications during somatic hypermutation: Implications for oncogenic translocations and heavy chain disease. Proc. Natl. Acad. Sci. USA 95, 2463-2468

Goyenechea, B., Klix, N., Yélamos, J., Williams, G. T., Riddell, A., Neuberger, M. S. and Milstein, C., *EMBO J.* 1997. *16*, 3987-3994

Green, N.S., Lin, M.M. and Scharff, M.D. (1998). Immunoglobulin hypermutation in cultured cells. Immunol. Rev. 162, 77-87.

Green, R. and Szostak, J.W. (1992) Science, 258, 1910-5.

Griffin, C. S., Simpson, P. J., Wilson, C. R. & Thacker, J. Mammalian recombination-repair genes XRCC2 and XRCC3 promote correct chromosome segregation. *Nat Cell Biol* **2**, 757-761. (2000)

Jain, R., Roncella, S., Hashimoto, S., Carbone, A., Franco di Celle, P., Foa, R., Ferrarini, M and Chiorazzi, N. (1994). A potential role for antigen selection in the clonal evolution of Burkitt's lymphoma. J. Immunol. 153, 45-52.

Johnson, R. D., Liu, N. & Jasin, M. Mammalian XRCC2 promotes the repair of DNA double-strand breaks by homologous recombination. *Nature* **401**, 397-399. (1999).

Jolly, C.J., Klix, N. and Neuberger, M.S. (1997). Rapid methods for the analysis of immunoglobulin gene hypermutation: application to transgenic and gene targeted mice. Nucleic Acids Res. 25, 1913-1919.

Jolly, C.J., Wagner, S.D., Rada, C.A., Klix, N., Milstein, C. and Neuberger, M.S. (1996). The targeting of somatic hypermutation. Semin. Immunol. 8, 159-168.

Joyce, G.F. (1994) *Curr. op. Structural Biol.*, 4, 331-336.

Kim, S., Humphries, E. H., Tjoelker, L., Carlson, L. & Thompson, C. B. Ongoing diversification of the rearranged immunoglobulin light-chain gene in a bursal lymphoma cell line. *Mol Cell Biol* **10**, 3224-3231. (1990)

Klein, G., Giovanella, B., Westman, A., Stehlin, J. and Mumford, D. (1975). An EBV negative cell line established from a American Burkitt lymphoma; receptor characteristics, EBV infectability and permanent conversion into EBV-positive sublines by in vitro infection. Intervirology 5, 319-334.

Klix *et al.*, (1998) Eur. J. Immunol. 28:317-326.

Liu, N., Lamerdin, J.E., Tebbs, R.S., Schild, D., Tucker, J.D., Shen, M.R., Brookman, K.W., Siciliano, M.J., Walter, C.A., Fan, W., Narayana, L.S., Zhou, Z.Q., Adamson, A.W., Sorenson, K.J., Chen, D.J., Jones, N.J. & Thompson, L.H. XRCC2 and XRCC3, new human Rad51-family members, promote chromosome stability and protect against DNA cross-links and other damages. *Mol Cell* **1**, 783-793. (1998).

Lundberg, K.S., *et al.*, (1991) Gene 108:1-6.

Luria and Delbreck., (1943) Genetics 28:491-511

Maizels, N. Somatic hypermutation: how many mechanisms diversify V region sequences? *Cell* **83**, 9-12. (1995).

Mattheakis, L.C., Bhatt, R.R. and Dower, W.J. (1994) Proc Natl Acad Sci U S A, 91, 9022-6.

Matthias, et al., (1989) NAR 17, 6418.

Max, E. E., Maizel, J. V. and Leder, P., *J. Biol. Chem.* 1981. *256*: 5116-5120.

McCafferty, J., Griffiths, A.D., Winter, G. and Chiswell, D.J. (1990) Nature, 348, 552-4.

McCormack, W. T., Hurley, E. A. & Thompson, C. B. Germ line maintenance of the pseudogene donor pool for somatic immunoglobulin gene conversion in chickens. *Mol Cell Biol* **13**, 821-830. (1993)

McKean, D., Huppi, K., Bell, M., Staudt, L., Gerhard, W. and Weigert, M. (1984). Generation of antibody diversity in the immune response of BALB/c mice to influenza virus haemagglutinin. Proc. Natl. Acad. Sci. USA 81, 3180-3184.

Moore, M.J. (1995) Nature, 374, 766-7.

Neuberger, M.S. and Milstein, C. (1995). Somatic hypermutation. Curr. Opin. Immunol. 7, 248-254.

Papavasiliou, F. N. & Schatz, D. G. Cell-cycle-regulated DNA double-stranded breaks in somatic hypermutation of immunoglobulin genes. *Nature* **408**, 216-221. (2000).

Parham, P. (ed). (1998). Somatic hypermutation of immunoglobulin genes. Immunological Reviews, Vol. 162 (Copenhagen, Denmark: Munksgaard).

Pierce, A. J., Johnson, R. D., Thompson, L. H. & Jasin, M. XRCC3 promotes omology-directed repair of DNA damage in mammalian cells. *Genes Dev* **13**, 2633-2638. (1999).

Rada, C., Ehrenstein, M. R., Neuberger, M. S. & Milstein, C. Hot spot focusing of somatic hypermutation in MSH2-deficient mice suggests two stages of mutational targeting. *Immunity* **9**, 135-141. (1998).

Ratech, H. (1992). Rapid cloning of rearranged immunoglobulin heavy chain genes from human B-cell lines using anchored polymerase chain reaction. Biochem. Biophys. Res. Commun. 182, 1260-1263

Reynaud, C. A., Anquez, V., Grimal, H. & Weill, J. C. A hyperconversion mechanism generates the chicken light chain preimmune repertoire. *Cell* **48**, 379-388. (1987).

Reynaud, C.A., Dahan, A., Anquez, V. & Weill, J.C. Somatic hyperconversion diversifies the single V_H gene of the chicken with a high incidence in the D region. *Cell* **59**, 171-183 (1989)

Rogozin, I.B. and Kolchanov, N.A. (1992). Somatic hypermutation of immunoglobulin genes. II. Influence of neighbouring base sequences on mutagenesis. Biochem. Biophys. Acta 1171, 11-18.

Rooney, C. M., A. B. Rickinson, D. J. Moss, G. M. Lenoir, and M. A. Epstein. 1984. Paired Epstein-Barr virus-carrying lymphoma and lymphoblastoid cell lines from Burkitt's lymphoma patients: comparative sensitivity to non-specific and to allo-specific cytotoxic responses in vitro. Int J Cancer. **34**:339-48

Rowe, M., D. T. Rowe, C. D. Gregory, L. S. Young, P. J. Farrell, H. Rupani, and A. B. Rickinson. 1987. Differences in B cell growth phenotype reflect novel patterns of Epstein-Barr virus latent gene expression in Burkitt's lymphoma cells. Embo J. **6**:2743-51

Sale, J. E. & Neuberger, M. S. TdT-accessible breaks are scattered over the immunoglobulin V domain in a constitutively hypermutating B cell line. *Immunity* **9**, 859-869. (1998).

Sharpe *et al.*, (1991) EMBO J. 10:2139-2145.

Smith, G.P. (1985) Science, 228, 1315-7.

Takata, M., Sasaki, M.S., Sonoda, E., Fukushima, T., Morrison, C., Albala, J.S., Swagemakers, S.M., Kanaar, R., Thompson, L.H. & Takeda, S. The Rad51 paralog Rad51B promotes homologous recombinational repair. *Mol Cell Biol* **20**, 6476-6482. (2000).

Takata, M., Sasaki, M.S., Sonoda, E., Morrison, C., Hashimoto, M., Utsumi, H., Yamaguchi-Iwai, Y., Shinohara, A. & Takeda, S. Homologous recombination and non-homologous end-joining pathways of DNA double-strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells. *Embo J* **17**, 5497-5508. (1998)

Takata, M., Sasaki, M.S., Tachiiri, S., Fukushima, T., Sonoda, E., Schild, D., Thompson, L.H. & Takeda, S. Chromosome Instability and Defective Recombinational Repair in Knockout Mutants of the Five Rad51 Paralogs. *Mol Cell Biol* **21**, 2858-2866. (2001).

Tomlinson, I.M. (1997). V Base database of human antibody genes. Medical Research Council, Centre for Protein Engineering, UK. http://www.mrc-cpe.cam.ac.uk/

Tuerk, C. and Gold, L. (1990) Science, 249, 505-10.

Wabl, M., Burrows, P.D., von Gabain, A. and Steinberg, C.M. (1985). Hypermutation at the immunoglobulin heavy chain locus in a pre-B-cell line. Proc. Natl. Acad. Sci. USA 82, 479-482.

Wagner, S.D., Milstein, C. and Neuberger, M.S. (1995). Codon bias targets mutation. Nature 376, 732.

Weill, J. C. & Reynaud, C. A. Rearrangement/hypermutation/gene conversion: when, where and why? *Immunol Today* **17**, 92-97. (1996)

Wilson, P.C., de Boutellier, O., Liu, Y-J., Potter, K., Banchereau, J., Capra, J.D. and Pascual, V. (1998). Somatic hypermutation introduces insertions and deletions into

immunoglobulin V genes. J. Exp. Med. 187, 59-70.

Wu, H., and Kaartinen, M., (1995) Scand. J. Immunol. 42:52-59.

Wu, H., Pelkonen, E., Knuutila, S and Kaartinen M. (1995). A human follicular lymphoma B cell line hypermutates its functional immunoglobulin genes in vitro. Eur. J. Immunol. 25, 3263-3269.

Yamaguchi-Iwai, Y., Sonoda, E., Buerstedde, J.M., Bezzubova, O., Morrison, C., Takata, M., Shinohara, A. & Takeda, S. Homologous recombination, but not DNA repair, is reduced in vertebrate cells deficient in RAD52. *Mol Cell Biol* **18**, 6430-6435. (1998).

Yélamos, J., Klix, N., Goyenechea, B., Lozano, F., Chui, Y.L., González Fernández, A., Pannell, R., Neuberger, M.S. & Milstein, C (1995). Targeting of non-Ig sequences in place of the V segment by somatic hypermutation. *Nature* **376**, 225-229.

Zhang, L., R. R. French, H. T. Chan, T. L. O'Keefe, M. S. Cragg, M. J. Power, and M. J. Glennie. 1995. The development of anti-CD79 monoclonal antibodies for treatment of B-cell neoplastic disease. Ther Immunol. **2**:191-202

Zhu, M., Rabinowitz, J.L., Green, N.S., Kobrin, B.J. and Scharff, M.D. (1995) A well-differentiated B cell line is permissive for somatic mutation of a transfected immunoglobulin heavy-chain gene. Proc. Natl. Acad. Sci. USA 92, 2810-2814.

Zou, X., Xian, J., Popov, A. V., Rosewell, I. R., Muller, M. and Brüggemann, M., *Eur. J. Immunol.* 1995. *25*: 2154-62.